A Systematic Review Evaluating Psychometric Properties of Parent/Caregiver report Instruments on Child Maltreatment — Part 1: Content Validity

Abstract

Aims: Child maltreatment (CM) is a serious public health issue, affecting over half of all children globally. Although most CM is perpetrated by parents or caregivers and their reports of CM is more accurate than professionals or children, parent or caregiver report instruments measuring CM have never been systematically evaluated for their content validity; the most important psychometric property. This systematic review aimed to evaluate the content validity of all current parent or caregiver report CM instruments.

Methods: A systematic literature search was performed in CINAHL, Embase, ERIC, PsycINFO, PubMed, and Sociological Abstracts; grey literature were retrieved through reference checking. Eligible studies needed to report on content validity of instruments measuring CM perpetrated and reported by parents or caregivers. The quality of studies and content validity of the instruments were evaluated using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guidelines.

Results: Fifteen studies reported on the content validity of fifteen identified instruments. The study quality was generally poor. The content validity of the instruments was overall sufficient, but most instruments did not provide high-quality evidence for content validity.

Conclusions: Most instruments included in this review showed promising content validity. The ISPCAN Child Abuse Screening Tool for use in Trial appears to be the most promising, followed by the Family Maltreatment-Child Abuse criteria. However, firm conclusions cannot be drawn due to the low quality of evidence for content validity. Further studies are required to evaluate the remaining psychometric properties for recommending parent or caregiver report CM instruments.

Keywords: Assessment; child abuse; COSMIN; measure; measurement properties; parent report

Introduction

Child maltreatment (CM) is defined by the World Health Organization (WHO, 2016) as:

> the abuse and neglect of children under 18 years of age. It includes all forms of physical and/or emotional ill treatment, sexual abuse, neglect, negligence and commercial or other exploitation, which results in actual or potential harm to the child's health, survival, development or dignity in the context of a relationship of responsibility, trust or power. (p. 94)

This broad definition can be distinguished into four subtypes of CM (Krug, Linda, James, Anthony, & Rafael, 2002; WHO, 1999): (1) physical abuse (PA: acts causing actual or potential physical harm); (2) emotional abuse (EA: acts having adverse impact on a child's emotional development); (3) sexual abuse (SA: acts using a child for sexual gratification); (4) neglect (failure in providing for the development of a child in health, education, emotional development, nutrition, shelter and safe living conditions).

CM causes significant public health problems and socioeconomic burden. CM can cause physical injuries, psychosocial difficulties, and lower academic achievement during childhood (Boden, Horwood, & Fergusson, 2007; Glaser, 2000; Teicher, Samson, Anderson, & Ohashi, 2016; van Harmelen et al., 2010). Moreover, adults with histories of childhood abuse tend to have higher risk of mortality, lower educational attainment, and lower income compared with adults without a history of CM (Anda, Butchart, Felitti, & Brown, 2010; Currie & Spatz Widom, 2010; Danese & McEwen, 2012; Felitti et al., 1998).

The prevalence of CM in the general population has been estimated at 57.6% of all children in the world (Hillis, Mercy, Amobi, & Kress, 2016) and most CM is perpetrated by

parents or caregivers (Devries et al., 2018; Sedlak et al., 2010). A recent meta-analysis on global prevalence of CM suggests that the overall prevalence rates are 12.7% for SA, 22.6% for PA, 36.3% for EA, 34.7% for neglect (Stoltenborgh, Bakermans-Kranenburg, Alink, & Ijzendoorn, 2015). While the most common perpetrators of SA are non-family members (Finkelhor, Shattuck, Turner, & Hamby, 2014), at least 50% of PA and EA or neglect is perpetrated by caregivers (Devries et al., 2018). For example, in the United States of America, parents are the perpetrators of 72% of all physically abused children, 73% of emotionally abused children, and 92% of neglected children, compared with 37% of sexually abused children (Sedlak et al., 2010). Thus, CM perpetrated by parents/caregivers is an important construct of interest.

However, estimates of the prevalence of CM vary markedly depending on who the informants are. Meta-analyses have shown that self-reported or caregiver-reported prevalence of CM is greater than prevalence reported by professionals, such as doctors or child protection workers (Stoltenborgh et al., 2015). Furthermore, the prevalence rate of most forms of CM, reported by children is far lower when compared with caregiver reports, with SA the notable exception (Devries et al., 2018). In contrast to self-report and caregiver-report, lower professional reported prevalence rates may be the result of professionals more likely to report severe CM cases, as mild cases may be considered as not important enough to report (Negriff, Schneiderman, & Trickett, 2017). Conversely, young children may have more trouble recalling abusive and neglecting behaviors than adult caregivers (Devries et al., 2018). While caregiver-reported prevalence on CM appears to be less affected by underestimation of CM (Devries et al., 2018; Stoltenborgh et al., 2015), accuracy and reliability of a caregiver-report instrument on CM are still an ongoing debate due to caregivers' general tendency to respond in socially desirable ways (Compier-de Block et al., 2017). Therefore, identifying reliable and valid parent or caregiver report measures is essential to estimate accurate prevalence of CM.

While directly measuring the prevalence of parental CM is important, there is a need to measure parents' attitude towards CM for the purpose of CM prevention, that is, parental values, beliefs, or feelings in relation to abusive and neglecting behavior towards a child (Altmann, 2008). Since parents are the main perpetrators of CM (Devries et al., 2018; Sedlak et al., 2010), prevention efforts need to focus on parents. Parents' attitude towards CM is a critical predictive factor of parental child abuse behavior (Stith et al., 2009). Several studies have shown that parents with more positive beliefs or values towards CM, tend to show more child abusive behaviors than parents with a negative attitude (Asadollahi, Jabraeili, Asghari Jafarabadi, & Hallaj, 2016; Ateah & Durrant, 2005; Bower-Russa, 2005; Chavis et al., 2013; Stith et al., 2009; Vittrup, Holden, & Buck, 2006). For this reason, a number of studies on CM prevention used instruments to measure parents' attitude towards CM as an outcome measure to establish whether the programs being evaluated are effective (Chen & Chan, 2016; Gershoff, Lee, & Durrant, 2017; Holden, Brown, Baldwin, & Croft Caderao, 2014; Voisine & Baker, 2012). Therefore, to measure the outcomes for evidence-based CM prevention programs, reliable and valid instruments to measure parents' attitude towards CM are needed, as well as suitable instruments to measure parents' actual maltreating behaviors towards their children.

Even though the selection of a high quality instrument is critically important for accurate and reliable assessment of CM, there is no universally accepted gold standard for measuring CM (Bailhache, Leroy, Pillet, & Salmi, 2013). The best way for selecting suitable evidence-based instruments is by evaluating the instruments' psychometric properties through a systematic review (Scholtes, Terwee, & Poolman, 2011). The Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) group has developed and published comprehensive guidelines for conducting systematic reviews on psychometric properties of patient-reported outcome instruments (Prinsen et al., 2018; Terwee, Prinsen, Chiarotto, Westerman, et al., 2018). The COSMIN methodological guidelines include a

taxonomy defining each psychometric property (Mokkink et al., 2010b); a checklist to assess the methodological quality of psychometric studies (Mokkink et al., 2018); criteria to evaluate the psychometric quality of instruments (Prinsen et al., 2018; Terwee, Prinsen, Chiarotto, Westerman, et al., 2018); and a rating system to summarize psychometric evidence and grade quality of all evidence used for the psychometric quality assessment of instruments (Prinsen et al., 2018; Terwee, Prinsen, Chiarotto, Westerman, et al., 2018).

The COSMIN taxonomy distinguishes nine psychometric properties across three domains: (1) validity (i.e., the extent to which an instrument measures the construct it is intended to measure); (2) reliability (i.e., the extent to which scores for patients who have not changed are the same for repeated measurements); and (3) responsiveness (i.e., the ability to detect clinically important change over time in the construct measured) (Mokkink et al., 2010b). The domain of validity contains five psychometric properties: content validity (i.e., the extent to which the content of an instrument adequately reflects the construct to be measured); structural validity (i.e., the extent to which the scores adequately reflect the dimensionality of the construct to be measured); cross-cultural validity (i.e., the extent to which a translated or culturally adapted version of an instrument adequately reflects the performance of the items of the original instrument); hypothesis testing for construct validity (i.e., the extent to which the scores are consistent with hypotheses on differences between relevant groups and relations to scores of other instruments); and criterion validity (i.e., the extent to which the scores adequately reflect a 'gold standard') (Mokkink et al., 2010b). Next, the reliability domain contains three psychometric properties: internal consistency (i.e., the degree of the interrelatedness of items); reliability (i.e., the proportion of total score variance which is due to true differences among respondents); and measurement error (i.e., the systematic and random error of a respondent's score that is not because of true changes in the construct measured) (Mokkink et al., 2010b). Lastly, the domain of responsiveness includes only one psychometric

property that is also called responsiveness, which has the same definition as the domain (Mokkink et al., 2010b).

When selecting an instrument, the most important psychometric property is its content validity (Prinsen et al., 2018; Prinsen et al., 2016); if it is unclear what construct(s) the instrument is actually measuring, then the evidence of the remaining psychometric properties is not valuable (Patrick et al., 2011; Streiner, Norman, & Cairney, 2015). For example, a high Cronbach's alpha does not guarantee that all important concepts are included. Similarly, a high test–retest reliability or adequate responsiveness does not imply that all items are relevant to the construct being measured (Cortina, 1993; Sijtsma, 2009).

Content validity pertains to three aspects of the content of an instrument: (1) relevance (i.e., the degree to which all items of an instrument are relevant for the construct of interest within a target population and purpose of use); (2) comprehensiveness (i.e., the degree to which all key concepts of the construct are included in an instrument); and (3) comprehensibility (i.e., the degree to which items of an instrument are easy to understand by respondents) (Terwee, Prinsen, Chiarotto, de Vet, et al., 2018). Weaknesses in any of these three aspects of content validity can impact on all other psychometric properties (Wiering, de Boer, & Delnoij, 2017) in the following ways: if items of an instrument are irrelevant (poor relevance), it may decrease interrelatedness among the items (internal consistency), structural validity, and interpretability of an instrument, and if an instrument misses some key concepts of the construct (poor comprehensiveness), it may reduce the ability of an instrument to detect real change in the construct of interest before and after intervention (poor responsiveness) (Terwee, Prinsen, Chiarotto, Westerman, et al., 2018). Since content validity can have a significant influence on all other psychometric properties, the COSMIN methodological guidelines recommend evaluating the content validity of an instrument first and to not evaluate other psychometric properties if reviewers have high-quality evidence that the instrument has

insufficient content validity (Prinsen et al., 2018; Terwee, Prinsen, Chiarotto, Westerman, et al., 2018).

To have good content validity, instrument items and instructions should be sufficiently relevant, comprehensive, and comprehensible, based on high-quality evidence (Chiarotto, 2019). According to the COSMIN criteria, for a measure to be rated as having good content validity the measure should have: (1) items relevant to the construct of interest in a specific population and purpose of use, and appropriate response options and a recall period (relevance); (2) comprehensive items covering all key concepts (comprehensiveness); and (3) instructions, items, and response options that are understandable to the target population (comprehensibility) (Terwee, Prinsen, Chiarotto, Westerman, et al., 2018). Evidence for rating these three aspects of content validity is mainly derived from instrument development and content validity studies (Terwee, Prinsen, Chiarotto, de Vet, et al., 2018). The development study refers to a study generating relevant items based on input from the target population for a new instrument (item generation) and evaluating comprehensiveness and comprehensibility of a draft instrument by interview or survey with the target population (cognitive interview/pilot test). The content validity study refers to a study asking target population and professionals about relevance, comprehensiveness, and comprehensibility of an existing instrument. As additional evidence, the original instrument (i.e., content of instrument itself) should also be rated based on subjective opinion of reviewers in terms of relevance, comprehensiveness, and comprehensibility (Terwee, Prinsen, Chiarotto, Westerman, et al., 2018). Summarizing all evidence from the studies and content of instrument itself, overall relevance, comprehensiveness, and comprehensibility of an instrument need to be determined (Terwee, Prinsen, Chiarotto, Westerman, et al., 2018). Furthermore, the level of quality of all evidence used to determine overall relevance, comprehensiveness, and comprehensibility should be summarised (graded) to show how confident we are in the overall ratings on the three aspects

of content validity respectively. When the overall relevance, comprehensiveness, and comprehensibility are all sufficient and the levels of quality of evidence for the overall ratings are all high, we can decisively conclude that the instruments has good content validity. (Terwee, Prinsen, Chiarotto, de Vet, et al., 2018).

Only one study to date has conducted a systematic review on content validity of CM instruments (Saini, Hoffmann, Pantelis, Everall, & Bousman, 2019). However, the review identified only child self-report and clinician interview instruments, which tend to underestimate the actual incidence of CM compared to parent report instruments (Devries et al., 2018), and one parent proxy-report instrument (asking parents about their children's maltreated experience by any adults, not about their own perpetration of CM) (Saini et al., 2019; Sprangers & Aaronson, 1992). None of the instruments and studies included in the review by Saini et al. (2019) overlapped with this current review for parent or caregiver reported CM instruments. Furthermore, the authors did not use the latest, thoroughly revised COSMIN methodological guidelines (Prinsen et al., 2018; Terwee, Prinsen, Chiarotto, Westerman, et al., 2018), but instead used the old version of the COSMIN checklist (Mokkink et al., 2010a) and criteria (Terwee et al., 2007) for assessing the methodological quality of studies on content validity and the quality of content validity of instruments. The old version of COSMIN checklist consists of a simplified 5-item for assessing only content validity studies and does not contains any standards for assessing the methodological quality of instrument development studies. Moreover, the early COSMIN criteria do not have specific consensus-based criteria for rating the relevance, comprehensiveness, and comprehensibility of an instrument (Terwee, Prinsen, Chiarotto, Westerman, et al., 2018). To address these shortcomings, the COSMIN methodological guideline for assessing content validity of an instrument has been recently developed to provide a detailed and standardised checklist and criteria (Terwee, Prinsen, Chiarotto, Westerman, et al., 2018). No other systematic reviews on

content validity or any of the other psychometric properties of parent or caregiver report instruments on CM have been published.

Study Aim

The aim of this systematic review was to evaluate content validity of all current parent or caregiver report CM instruments using the updated COSMIN methodological guidelines (Prinsen et al., 2018; Terwee, Prinsen, Chiarotto, Westerman, et al., 2018). Due to the size, scope and complexity of reporting the remaining psychometric properties, we aim to report the quality of studies and psychometrics of instruments identified in this systematic review in a companion paper (Part 2), excluding those instruments found to have high-quality evidence for insufficient content validity in this paper.

Methods

This systematic review was conducted and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Moher, Liberati, Tetzlaff, Altman, & PRISMA Group, 2009) and the COSMIN methodological guidelines (Prinsen et al., 2018; Terwee, Prinsen, Chiarotto, Westerman, et al., 2018). This review consists of three consecutive steps (see Figure 1):

- Step 1: *Systematic literature search* formulating eligibility criteria (step 1.1), and searching literatures and selecting studies (step 1.2) (Moher et al., 2009);
- Step 2: *Evaluation of the methodological quality of studies* on instrument development (step 2.1) and content validity (step 2.2) using the COSMIN Risk of Bias checklist (Mokkink et al., 2018);
- Step 3: *Evaluation of the content validity of instruments* rating the result of single studies against the criteria for good content validity (step 3.1), summarizing all results of studies per instrument (step 3.2), and grading quality of evidence on content validity (step 3.3) (Terwee, Prinsen, Chiarotto, Westerman, et al., 2018).

Each of these steps will be explained more detail in the following sections.

*** Insert Figure 1 about here***

Systematic Literature Search (Step 1)

The systematic literature search was conducted for both this paper on content validity (Part 1) and a companion paper on other psychometric properties (Part 2) by formulating eligibility criteria (step 1.1), and searching literature and selecting studies (step 1.2).

Eligibility criteria (step 1.1). To select instruments and studies for this current review, the following five eligibility criteria for inclusion were used: (1) parent or caregiver report instruments assessed their own attitudes toward CM or maltreating behaviors towards their children; (2) at least one subscale or a minimum of 30% of all items within an instrument referred to one or more types of CM (i.e., PA, EA, SA, and neglect) (Krug et al., 2002; WHO, 1999), as a criterion to ensure the contribution to the overarching construct of an instrument was involved CM; (3) instruments were developed and studies were published in English; (4) studies reported on psychometric data of at least one of the nine psychometric properties of eligible instruments as defined in the COSMIN taxonomy (Mokkink et al., 2010b) that were published as original journal articles, manuals, book chapters or conference papers; and (5) studies on content validity reported on the development of new items of eligible instruments, and/or evaluated the relevance, comprehensiveness, or comprehensibility of the content of the eligible instruments as reported by parents/caregivers and/or professionals.

Literature search and study selection (step 1.2). To identify eligible instruments and journal articles that reported on any psychometric properties of the instruments as defined in the COSMIN taxonomy (Mokkink et al., 2010b), systematic literature searches were performed in six electronic databases (CINAHL, Embase, ERIC, PsycINFO, PubMed, and Sociological Abstracts) on the 29th of January 2018 with an update on the 5th of October 2019. Search terms consisted of subject headings and free text words (see Online Appendix A). All publications

prior to October 2019 were considered for inclusion.

Abstracts and articles retrieved from database searches were screened to identify eligible instruments and journal articles on any psychometric property by two reviewers independently. One reviewer screened all abstracts while the other reviewer screened a random selection of approximately half of all abstracts; all full texts of eligible abstracts were retrieved and screened by both independent reviewers. Any discrepancies between both reviewers were resolved by involving a third reviewer. The degree of agreement between the two reviewers was assessed using Cohen's weighted Kappa (Cohen & Humphreys, 1968); agreement was very good (Altman, 1991): (1) Weighted Kappa for abstract selection = 0.87 (95% confidence interval [CI] = [0.83, 0.90]); (2) Weighted Kappa for article selection = 0.86 (95% CI [0.77, 0.94]).

Reference lists of all included full-text articles on any psychometric property were hand searched to identify additional eligible instruments and psychometric studies on the instruments. Websites of Pearson and Western Psychological Services, two major measurement publishers in social science, were also searched to retrieve potential instruments and manuals not identified in previous databases and reference searches. Both of the reference lists and websites were searched by one reviewer and the additionally retrieved instruments and psychometric studies were checked by another reviewer. If instruments were not published or freely available, the developers of the instruments were contacted by e-mail to retrieve the original instruments.

Finally, among all eligible psychometric studies, only studies on content validity (i.e., instrument development and content validity studies) were included in this review (Part 1) for the evaluation of content validity. Studies on other psychometric properties were excluded in this paper (Part 1), as these findings will be reported on in a companion paper (Part 2).

Evaluation of Methodological Quality of Studies (Step 2)

The methodological quality of included studies on instrument development (step 2.1) and content validity (step 2.2) was assessed using the COSMIN Risk of Bias checklist (Mokkink et al., 2018). First, the development studies were assessed using 35 items from the checklist, which consists of a separate rating of the quality of the 'instrument design' (item generation) to ensure relevance of a new instrument and 'cognitive interview/pilot test' to evaluate comprehensiveness and comprehensibility of a draft instrument (Terwee, Prinsen, Chiarotto, de Vet, et al., 2018). Next, content validity studies were assessed using 38 items from the checklist, comprised of one set of items assessing quality of studies that ask parents/caregivers about relevance, comprehensiveness, and comprehensibility, and another set assessing quality of studies that ask professionals about relevance and comprehensiveness (Terwee, Prinsen, Chiarotto, de Vet, et al., 2018). Total ratings for each aspect of content validity (i.e., relevance, comprehensiveness, and comprehensibility) were determined separately. Separate total ratings were also determined for the two parts of the development study ('instrument design' and 'cognitive interview/pilot test'), as well as for two types of content validity study ('asking parents/caregivers' and 'asking professionals') (Mokkink et al., 2018).

When rating the methodological quality of the instrument development and content validity studies, each checklist item was ranked on a four-point rating scale (1 = inadequate, 2 = doubtful, 3 = adequate, and 4 = very good). A total rating for relevance, comprehensiveness, or comprehensibility was obtained by calculating a percentage of the ratings based on the following formula (Cordier et al., 2015), instead of a worst score counts method (reporting total ratings gained by taking the lowest rating among any of the checklist items) recommend by the COSMIN methodological guidelines (Mokkink et al., 2018). This approach was adopted as determining total scores of methodological quality of studies that are entirely based on the lowest rating of single items impedes the detection of subtle differences in methodological

quality between studies (Speyer, Cordier, Kertscher, & Heijnen, 2014).

$$\textbf{Total score for methodological quality} = \frac{(\textbf{Total score obtained} - \textbf{Min score possible})}{(\textbf{Max score possible} - \textbf{Min score possible})} \times \textbf{100}\%.$$

The total percentage score is then categorised into the following four scores: inadequate (from 0 to 25%), doubtful (from 25.1 to 50%), adequate (from 50.1 to 75%), and very good (from 75.1 to 100%). Two reviewers rated the methodological quality independently where after consensus ratings were determined between the two reviewers. The inter-rater reliability was calculated using weighted Kappa (Cohen & Humphreys, 1968) between both reviewers.

After assessment of methodological quality on the included instrument development and content validity studies, the following data were extracted from the included studies and instruments: (1) study characteristics (i.e., study purpose, study population, and parents or professionals involvement); (2) instrument characteristics (i.e., instrument names and acronyms, measured constructs, targeted population, purpose of use, number of [sub] scales, number of items, response options and recall period); and (3) study results on all three aspects of content validity (relevance, comprehensiveness, and comprehensibility). All relevant data were extracted by one reviewer and rechecked for accuracy by another reviewer.

Evaluation of Content Validity of Instruments (Step 3)

The content validity of instruments was assessed for three separate aspects of content validity (relevance, comprehensiveness, and comprehensibility) in three sequential steps: step 3.1, step 3.2, and step 3.3. All ratings were conducted by two reviewers independently and any discrepancies were resolved by consensus.

Rating the result of single studies (step 3.1). Rating the results of single studies was conducted for each instrument development study, content validity study, and content of the instrument itself separately. The results of each development and content validity study was rated based on the qualitative or quantitative data obtained by asking parents/caregivers or

professionals about content validity of an instrument, using the ten pre-defined criteria on relevance (5), comprehensiveness (1), and comprehensibility (4) (Terwee, Prinsen, Chiarotto, Westerman, et al., 2018). By using the same criteria, the content of the original instrument itself (items, response options, and recall period) was also rated based on the subjective judgement of the reviewers. The reviewers received extensive training in appraising content validity of instruments using the COMSIN criteria under supervision of the second author who has considerable expertise in psychometrics and the COSMIN framework. Ratings for each source of evidence on content validity were given as sufficient (85% or more of the instrument items meet the criterion: +), insufficient (less than 85% of the instrument items meet the criterion: -), or indeterminate (lack of evidence to determine the quality or inadequate methodological quality of studies: ?). More detailed information on these criteria and how to apply these criteria can be found in the user manual on COSMIN methodology for assessing content validity (Terwee, Prinsen, Chiarotto, de Vet, et al., 2018).

Summarizing the results of all studies per instrument (step 3.2). All results from available studies on development and content validity per instrument and the reviewers' ratings on content of the instrument were qualitatively summarized into overall ratings for relevance, comprehensiveness, and comprehensibility of the instrument (i.e., all ratings determined in the previous step were jointly assessed) (Terwee, Prinsen, Chiarotto, de Vet, et al., 2018; Terwee, Prinsen, Chiarotto, Westerman, et al., 2018). The focus in this step was on the specific instrument, while in the previous step the focus was on single studies. An overall sufficient (+), insufficient (-) inconsistent (±), or indeterminate (?) rating was given for relevance, comprehensiveness, and comprehensibility for each instrument (Terwee, Prinsen, Chiarotto, de Vet, et al., 2018; Terwee, Prinsen, Chiarotto, Westerman, et al., 2018). For example, if all relevance scores of development studies, content validity studies, and content of the instrument (reviewers' ratings) were sufficient, insufficient, or indeterminate, the overall relevance rating

became sufficient (+), insufficient (-), or indeterminate (?). If, however, at least one of these three scores was inconsistent with the other two scores, the overall rating became inconsistent (±). An exception to this rule was when the scores of both development and content validity studies were all indeterminate and inconsistent with the reviewers' rating on content of the instrument. In this instance, the overall rating could be determined by solely the reviewers' rating. Further details on rating overall relevance, comprehensiveness, and comprehensibility can be founded in the user manual for assessing content validity (Terwee, Prinsen, Chiarotto, de Vet, et al., 2018).

Grading the quality of evidence on content validity (step 3.3). The quality of the evidence (i.e., the total body of evidence used for overall ratings on relevance, comprehensiveness and comprehensibility of an instrument) was graded (high, moderate, low, or very low) using a modified Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach (Guyatt et al., 2008; Terwee, Prinsen, Chiarotto, de Vet, et al., 2018; Terwee, Prinsen, Chiarotto, Westerman, et al., 2018). The GRADE approach is used to downgrade level of evidence when there are concerns about the quality of evidence. The starting point of the evidence quality rating is based on the assumption that the overall rating is of high quality. Next, ratings are downgraded one or more levels (to moderate, low or very low) if there is serious or very serious risk of bias (i.e., limitations in the methodological quality of studies), inconsistency (i.e., unexplained heterogeneity in results of studies), and/or indirectness (i.e., evidence from different populations than the target population of interest in the review) (Terwee, Prinsen, Chiarotto, de Vet, et al., 2018; Terwee, Prinsen, Chiarotto, Westerman, et al., 2018). The quality of evidence was not graded if the overall rating was indeterminate (?) due to lack of evidence. More specific information about grading the quality of evidence can be found in the COSMIN user manual for content validity (Terwee, Prinsen, Chiarotto, de Vet, et al., 2018).

Results

Systematic Literature Searches

In total, 2,859 non-duplicate abstracts were identified from six databases: CINAHL (1,173 records); Embase (456 records); ERIC (523 records); PsycINFO (285 records); PubMed (1,092 records); Sociological Abstracts (133 records). Figure 2 shows the flow diagram of the studies and instruments identified during the literature search and screening process in accordance with PRISMA (Moher et al., 2009). A total of 253 full-text articles and 164 instruments were assessed for eligibility, resulting in 23 full-text articles reporting on psychometric properties and fourteen instruments. Online Appendix B summarizes a list of the 150 excluded instruments and reasons for exclusion.

Insult Figure 2 about here

Reference checking of the 23 articles on psychometric properties resulted in one additional instrument and ten additional psychometric studies being identified as meeting eligibility criteria. A total of 33 psychometric studies evaluating 15 different instruments were identified. Fifteen out of 33 psychometric studies reported on content validity (i.e., instrument development or content validity studies) and were included in this review (Part 1).

Characteristics of Included Studies and Instruments

Descriptions of the instrument development or content validity studies of the included CM instruments are presented in Online Appendix C. Table 1 provides a summary of the characteristics of all fifteen instruments, including names and acronyms, construct of interest (subscales), target population, intended contexts for use, number of (sub)scales and items, response options, and recall periods. All fifteen instruments measured at least one type of CM (construct of interest) for parents or caregivers (target population) with the purpose to identify maltreating parents, as well as abused children, and/or to evaluate intervention programs (purpose of use). Of the fifteen instruments identified, no instrument measured only SA; three

measured both SA and other types of CM (PA, EA, and/or neglect); and twelve measured other types of CM. The total number of subscales ranged from no subscales to six subscales; the total number of items varied between 4 and 60. All but one instrument used a Likert response scale, while only the P-CAAM used a reaction time response. Recall period varied between last week and last year for eight instruments (CNQ, CNS-MMS, CTSPC, FM-CA, ICAST-Trial, MCNS, MCNS-SF, and PRCM); the recall period was unspecified in the remaining seven instruments.

Insert Table 1 about here

Methodological Quality of Development and Content Validity Studies

The methodological quality of the fifteen included studies on instrument development (14) and content validity (10) was assessed using the COSMIN checklist (Mokkink et al., 2018). All ten content validity studies overlapped with the development studies; one study reported on more than one instrument. An overview of all methodological quality ratings is presented in Table 2. Only five development studies reported on either item generation or cognitive interviewing. Of those five studies, three studies used both item generation and cognitive interviews, whereas the other two studies conducted cognitive interviews only. Of the thirteen instrument development study quality ratings, a single rating for relevance and comprehensiveness was classified as doubtful, while all other eleven ratings were classified as inadequate. In content validity studies, all but five studies asked parents/carers or professionals about at least one of the three aspects on content validity (relevance, comprehensiveness, and comprehensibility). Of the fifteen content validity study quality ratings, only three ratings (one relevance and two comprehensibility) were rated as very good or adequate, whereas all other twelve ratings were rated as doubtful or inadequate. No information was retrieved on comprehensiveness in any content validity studies. The inter-rater reliability for study quality assessment between both reviewers was good (weighted Kappa 0.76; 95% CI [0.68, 0.85]).

*** Insert Table 2 about here***

Content Validity of Instruments

Table 3 summarizes ratings on the content validity for development and content validity studies respectively, as well as the content of instrument itself involving fifteen studies and fifteen instruments. The data of each single study and content of instruments were evaluated against the ten criteria for good content validity for the following three separate aspects of content validity: relevance, comprehensiveness and comprehensibility (Terwee, Prinsen, Chiarotto, de Vet, et al., 2018; Terwee, Prinsen, Chiarotto, Westerman, et al., 2018). All development and content validity studies received indeterminate ratings, except for the following two studies of FM-CA: one development study received sufficient rating in relevance; and one content validity study received sufficient rating in comprehensibility. All but four instruments (CTS-ES, P-CAAM, POQ, and PRCM) were rated as sufficient for content of instruments based on the reviewers' expert opinion. Three instruments reported conflicting ratings in one of the three aspects of content validity (CTS-ES and POQ in relevance and PRCM in comprehensibility). Two instruments reported insufficient ratings in comprehensiveness (CTS-ES and POQ) and one instrument reported indeterminate ratings in all three aspects (P-CAAM).

*** Insert Table 3 about here***

Table 4 presents the overall ratings on content validity with quality of evidence for content validity. All but four instruments (CTS-ES, P-CAAM, POQ, and PRCM) received sufficient overall ratings in all three aspects of content validity (relevance, comprehensiveness, comprehensibility). Three instruments reported conflicting overall ratings in one of the three aspects of content validity (CTS-ES and POQ in relevance and PRCM in comprehensibility). Two instruments reported insufficient overall ratings in comprehensiveness (CTS-ES and POQ) and one instrument reported indeterminate overall ratings in all three aspects due to failure of retrieving the original instrument (P-CAAM).

*** Insert Table 4 about here***

High-quality evidence supporting overall ratings on content validity was only available for the FM-CA and the ICAST-Trial, whereas no high-quality evidence for content validity was found for the remaining thirteen instruments. In fact, 67% (30/45) of all evidence quality ratings for content validity were rated as very low. For overall ratings of relevance, six instruments received very low-quality of evidence ratings (APT, CNS-MMS, CTSPC, MCNS, MCNS-SF, PRCM). Three instruments were rated as having low-quality of evidence (CTS-ES, POQ, and SBS-SV); four instruments were rated as having moderate-quality of evidence (AAPI-2, CNQ, FM-CA, and IPPS); one instrument (ICAST-Trial) was rated as having high-quality of evidence; and one instrument (P-CAAM) was not evaluated (NE) because of indeterminate overall ratings (i.e., lack of evidence). All instruments received a very low-quality of evidence for the overall ratings in comprehensiveness, except for the following two instruments: CTSPC reported low-quality evidence and P-CAAM was not evaluated (NE). For overall ratings of comprehensibility, only two instruments received high-quality of evidence ratings (FM-CA and ICATS-Trial), whereas all other instruments (except CTSPC and P-CAAM) received very low ratings.

Discussion

The aim of this systematic review was to determine the quality of content validity of all current parent/caregiver report instruments measuring child maltreatment by parents or caregivers. This review identified fifteen instruments and fifteen corresponding instrument development and content validity studies of the instruments. Findings from the systematic review demonstrate lack of high-quality evidence, suggesting that none of the instruments received high quality ratings for all three aspects of content validity (relevance, comprehensiveness, and comprehensibility). As such, none of the instruments have unequivocally support for their use in terms of the quality of content validity.

Instrument Development Study

The majority of instrument development studies did not address SA as a construct of interest to be measured. While most CM instruments had a scale or subscale related to PA, EA, and/or neglect, only three instruments had some items or a subscale related to SA: a single item of the CTS-ES, two items of the ICAST-Trial, and one optional supplementary subscale of the CTSPC. A recent meta-analysis on who perpetrates CM reported that most SA is perpetrated by people other than parents/caregivers compared with the other three types of CM, but this result was only based on child self-report and professional report instruments due to lack of studies reporting SA by using parent report instruments (Devries et al., 2018). To verify the exceptional lower prevalence rates of SA perpetrated by parents, comparison of prevalence rates reported by parents, children, and professionals should be conducted. However, based on the findings from this review, comparing the prevalence rates of SA reported between parents/caregivers, children and professionals may be challenging because of the lack of parent report instruments on SA.

Many instrument development studies generated new items without involvement of the target population (parents/caregivers), that is, most instrument items were generated based on a review of relevant literature, commonly used instruments, or professional input by developers themselves. Involvement of the target population is essential to ensure adequate content validity in the generation of new instrument items (Terwee, Prinsen, Chiarotto, Westerman, et al., 2018). Involving the target population through individual interviews or focus groups helps to identify items that are relevant to the target population, to ensure items are based on their own experience or perceptions related to the construct being measured (Ricci et al., 2018). If the respondents (target population) are of the opinion that the instrument items are irrelevant, the instrument could fail to measure respondents' attitudes and behaviors accurately (Wiering et al., 2017). Therefore, development studies of new instrument items as reported in this review

may have significant methodological flaws given the lack of target population involvement.

Content Validity Study

Only a few content validity studies asked parents/caregivers about relevance, comprehensiveness, and comprehensibility of the instruments and reported specific research methods and results, which enabled the evaluation of the content validity of the instruments clearly. According to findings on the methodological quality of content validity studies, relevance of the final version of instruments was mostly evaluated by asking the professionals, whereas, surprisingly, the comprehensiveness of instruments was not evaluated by neither professionals nor parents/caregivers. Furthermore, the comprehensibility (i.e., how easy it is for respondents to understand instrument items) was rarely evaluated by parents/caregivers as respondents. The few studies that did evaluate the relevance and comprehensibility of instruments using parents/caregivers as respondents, lacked the required detail when reporting on the methodology (e.g., insufficient reporting on study design and results). This weaknesses made it difficult to determine if the content validity of instruments was positive or negative based on the evidence obtained from the content validity studies.

Synthesis of Evidence on Content Validity

Given that content validity is the first psychometric property to consider when selecting an instrument, the inadequate quality of evidence on content validity make it difficult to select the best instrument(s) (Terwee, Prinsen, Chiarotto, Westerman, et al., 2018). The majority of ratings (88/99) on relevance, comprehensiveness, and comprehensibility based on the development and content validity studies were categorised as indeterminate. Due to these indeterminate study ratings, most overall ratings on relevance, comprehensiveness, and comprehensibility were determined based on reviewers' subjective opinion about the content of instrument itself only. The results indicate lack of evidence on content validity or inappropriate methodological approaches used for instrument development and content validity studies

(Terwee, Prinsen, Chiarotto, Westerman, et al., 2018). Due to the largely inappropriate methodological approaches used when developing new instruments and assessing content validity of the instruments, in most instances, evidence on the quality of relevance, comprehensiveness, and comprehensibility was very low; high quality evidence was found only for the relevance or comprehensibility for two instruments(FM-CA and ICAST-Trial). Therefore, findings from this review indicate that evidence of the quality of content validity of parent/caregiver report CM instruments is very uncertain.

Based on available evidence on content validity for the fifteen included instruments, the ICAST-Trial seems to be the most promising instrument in terms of content validity; however, the evidence is not conclusive. The ICAST-Trial displayed high-quality evidence for sufficient relevance and comprehensibility, and very low evidence for sufficient comprehensiveness. The next most promising instrument was the FM-CA with high-quality evidence for sufficient comprehensibility, moderate evidence for sufficient relevance, and very low evidence for sufficient comprehensiveness. While none of the remaining thirteen instruments reported high-quality evidence on any aspects of content validity, they also have the potential to be used in terms of content validity, because no high-quality evidence for insufficient relevance, comprehensiveness, or comprehensibility was found.

Recommendation for Future Research

Based on the preceding discussion, there is a need for follow-up studies on parent-reported CM questionnaires to be conducted with the following five recommendations in mind. First, future instrument development studies should include SA parent reported items or subscales, especially in the case of early childhood sexual abuse where recall bias in young children is an important consideration. Second, development of a new instrument items should involve parents/caregivers (e.g., individual or group interviews) to identify relevant items from their perspective on CM. Third, additional validation studies are needed to evaluate content

validity of the included instruments, as current evidence on their content validity is not enough to determine conclusively which of the instruments has good content validity. In particular, the comprehensibility of the instruments should be further evaluated from the perspectives of parents/caregivers. Fourth, it is recommended that future studies apply the COSMIN guidelines in their study design for the generation of new items and assessment of content validity of instruments. Finally, a review on quality of the remaining psychometric properties of current parent/caregiver report CM instruments is needed, as no high-quality evidence of insufficient content validity was found. This additional assessment of psychometric quality will help clinicians and researchers decided which instruments to use for their interventions and research on CM perpetrated by parents/caregivers.

Limitations

This systematic review has some limitations. Firstly, only instruments developed and validated in English and psychometric studies published in English were considered. Thus, findings on content validity of parent/carer report CM instruments developed in languages other than English may have been excluded. Secondly, despite contacting the developer of the P-CAAM, we failed to retrieve the original instrument from the authors or from literature and, therefore, could not determine the overall ratings on content validity of this instrument. Lastly, while rating the quality of the studies and psychometric properties using the COSMIN guidelines for assessing content validity required a degree of subjective judgment by reviewers, all ratings for this review were conducted by two reviewers independently and disagreements were resolved through consensus.

Conclusion

Fifteen parent/caregiver report CM instruments were retrieved. An evaluation of the content validity using the COSMIN methodological guidelines found that the ICAST-Trial appear to be the most promising instrument, followed by the FM-CA, but firm conclusions

cannot be drawn because evidence concerning the content validity is limited and mostly of low quality. However, no high-quality evidence was found to indicate that the content validity is insufficient. As such, all identified instruments have the potential to be used, but their remaining psychometric properties should be evaluated. A companion paper (Part 2) will report on the evaluation of the remaining psychometric properties of the thirteen included instruments to identify parent/caregiver report instruments of CM with robust psychometric properties based on current evidence.

References

Altman, D. G. (1991). *Practical statistics for medical research*. London: Chapman and Hall.

Altmann, T. K. (2008). Attitude: A Concept Analysis. *Nursing Forum, 43*(3), 144-150. doi:10.1111/j.1744-6198.2008.00106.x

Anda, R. F., Butchart, A., Felitti, V. J., & Brown, D. W. (2010). Building a Framework for Global Surveillance of the Public Health Implications of Adverse Childhood Experiences. *American Journal of Preventive Medicine, 39*(1), 93-98. doi:10.1016/j.amepre.2010.03.015

Asadollahi, M., Jabraeili, M., Asghari Jafarabadi, M., & Hallaj, M. (2016). Parents' Attitude Toward Child Abuse Conducted in the Health Centers of Tabriz. *International Journal of School Health, 3*(3), e60221. doi:10.17795/intjsh-31198

Ateah, C. A., & Durrant, J. E. (2005). Maternal use of physical punishment in response to child misbehavior: implications for child abuse prevention. *Child Abuse & Neglect, 29*(2), 169-185. doi:10.1016/j.chiabu.2004.10.010

Bailhache, M., Leroy, V., Pillet, P., & Salmi, L. R. (2013). Is early detection of abused children possible?: a systematic review of the diagnostic accuracy of the identification of abused children. *BMC Pediatr, 13*, 202. doi:10.1186/1471-2431-13-202

Bavolek, S. J., & Keene, R. G. (1999). *Adult-Adolescent Parenting Inventory-AAPI-2: Administration and development handbook*. Park City, UT: Family Development Resources, Inc.

Bavolek, S. J., Kline, D. F., McLaughlin, J. A., & Publicover, P. R. (1979). Primary prevention of child abuse and neglect: Identification of high-risk adolescents. *Child Abuse & Neglect, 3*(3), 1071-1080. doi:10.1016/0145-2134(79)90152-2

Boden, J. M., Horwood, L. J., & Fergusson, D. M. (2007). Exposure to childhood sexual and physical abuse and subsequent educational achievement outcomes. *Child Abuse & Neglect, 31*(10), 1101-1114. doi:10.1016/j.chiabu.2007.03.022

Bower-Russa, M. (2005). Attitudes Mediate the Association Between Childhood Disciplinary History and Disciplinary Responses. *Child Maltreatment, 10*(3), 272-282. doi:10.1177/1077559505277531

Chavis, A., Hudnut-Beumler, J., Webb, M. W., Neely, J. A., Bickman, L., Dietrich, M. S., & Scholer, S. J. (2013). A brief intervention affects parents' attitudes toward using less physical punishment. *Child Abuse & Neglect, 37*(12), 1192-1201. doi:10.1016/j.chiabu.2013.06.003

Chen, M., & Chan, K. L. (2016). Effects of Parenting Programs on Child Maltreatment Prevention: A Meta-Analysis. *Trauma, Violence, & Abuse, 17*(1), 88-104. doi:10.1177/1524838014566718

Chiarotto, A. (2019). Patient-Reported Outcome Measures: Best Is the Enemy of Good (But What if Good Is Not Good Enough?). *Journal of Orthopaedic & Sports Physical Therapy, 49*(2), 39-42. doi:10.2519/jospt.2019.0602

Cohen, J., & Humphreys, L. H. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin, 70*(4), 213-220. doi:10.1037/h0026256

Compier-de Block, L. H. C. G., Alink, L. R. A., Linting, M., van den Berg, L. J. M., Elzinga, B. M., Voorthuis, A., . . . Bakermans-Kranenburg, M. J. (2017). Parent-Child Agreement on Parent-to-Child Maltreatment. *Journal of Family Violence, 32*(2), 207-217. doi:10.1007/s10896-016-9902-3

Cordier, R., Speyer, R., Chen, Y., Wilkes-Gillan, S., Brown, T., Bourke-Taylor, H., . . . Leicht, A. (2015). Evaluating the Psychometric Quality of Social Skills Measures: A

Systematic Review. *PLoS One, 10*(7), e0132299-e0132299. doi:10.1371/journal.pone.0132299

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98–104. doi:10.1037/0021-9010.78.1.98

Currie, J., & Spatz Widom, C. (2010). Long-Term Consequences of Child Abuse and Neglect on Adult Economic Well-Being. *Child Maltreatment, 15*(2), 111-120. doi:10.1177/1077559509355316

Danese, A., & McEwen, B. S. (2012). Adverse childhood experiences, allostasis, allostatic load, and age-related disease. *Physiology & Behavior, 106*(1), 29-39. doi:10.1016/j.physbeh.2011.08.019

Devries, K., Knight, L., Petzold, M., Merrill, K. G., Maxwell, L., Williams, A., . . . Abrahams, N. (2018). Who perpetrates violence against children? A systematic analysis of age-specific and sex-specific data. *BMJ Paediatrics Open, 2*(1). doi:10.1136/bmjpo-2017-000180

Felitti, V. J., Anda, R. F., Nordenberg, D., Williamson, D. F., Spitz, A. M., Edwards, V., . . . Marks, J. S. (1998). Relationship of Childhood Abuse and Household Dysfunction to Many of the Leading Causes of Death in Adults: The Adverse Childhood Experiences (ACE) Study. *American Journal of Preventive Medicine, 14*(4), 245-258. doi:10.1016/S0749-3797(98)00017-8

Finkelhor, D., Shattuck, A., Turner, H. A., & Hamby, S. L. (2014). The lifetime prevalence of child sexual abuse and sexual assault assessed in late adolescence. *Journal of Adolescent Health, 55*(3), 329-333. doi:10.1016/j.jadohealth.2013.12.026

Gershoff, E. T., Lee, S. J., & Durrant, J. E. (2017). Promising intervention strategies to reduce parents' use of physical punishment. *Child Abuse & Neglect, 71*, 9-23. doi:10.1016/j.chiabu.2017.01.017

Glaser, D. (2000). Child Abuse and Neglect and the Brain—A Review. *Journal of Child Psychology and Psychiatry, 41*(1), 97-116. doi:10.1111/1469-7610.00551

Gordon, D. A., Jones, R. H., & Nowicki, S. (1979). A Measure of Intensity of Parental Punishment. *Journal of Personality Assessment, 43*(5), 485-496. doi:10.1207/s15327752jpa4305_9

Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., . . . Grade Working Group. (2008). GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ, 336*(7650), 924-926. doi:10.1136/bmj.39489.470347.AD

Heyman, R. E., Snarr, J. D., Slep, A. M. S., Baucom, K. J. W., & Linkh, D. J. (2019). Self-reporting DSM–5/ICD-11 clinically significant intimate partner violence and child abuse: Convergent and response process validity. *Journal of Family Psychology,* Advance online publication. doi:10.1037/fam0000560

Hillis, S., Mercy, J., Amobi, A., & Kress, H. (2016). Global Prevalence of Past-year Violence Against Children: A Systematic Review and Minimum Estimates. *Pediatrics, 137*(3). doi:10.1542/peds.2015-4079

Holden, G. W., Brown, A. S., Baldwin, A. S., & Croft Caderao, K. (2014). Research findings can change attitudes about corporal punishment. *Child Abuse & Neglect, 38*(5), 902-908. doi:10.1016/j.chiabu.2013.10.013

Holden, G. W., & Zambarano, R. J. (1992). Passing the rod: Similarities between parents and their young children in orientations toward physical punishment. In I. E. Sigel, A. V. McGillicuddy-DeLisi, & J. J. Goodnow (Eds), *Parental belief systems: The psychological consequences for children* (2nd ed., pp. 143-172). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Kirisci, L., Dunn, M. G., Mezzich, A. C., & Tarter, R. E. (2001). Impact of Parental Substance Use Disorder and Child Neglect Severity on Substance Use Involvement in Male Offspring. *Prevention Science, 2*(4), 241-255. doi:10.1023/a:1013662132189

Krug, E. G., Linda, L. D., James, A. M., Anthony, B. Z., & Rafael, L. (Eds.). (2002). *World report on violence and health*. Geneva, Switzerland: Word Healh Organization.

Lang, J. M., & Connell, C. M. (2017). Development and validation of a brief trauma screening measure for children: The Child Trauma Screen. *Psychological Trauma: Theory, Research, Practice, and Policy, 9*(3), 390-398. doi:10.1037/tra0000235

Loeber, R., Farrington, D. P., Stouthamer-Loeber, M., & Van Kammen, W. B. (1998). *Antisocial behavior and mental health problems: Explanatory factors in childhood and adolescence*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Lounds, J. J., Borkowski, J. G., & Whitman, T. L. (2004). Reliability and Validity of the Mother-Child Neglect Scale. *Child Maltreatment, 9*(4), 371-381. doi:10.1177/1077559504269536

Meinck, F., Boyes, M. E., Cluver, L., Ward, C. L., Schmidt, P., Destone, S., & Dunne, M. P. (2018). Adaptation and psychometric properties of the ISPCAN Child Abuse Screening Tool for use in trials (ICAST-Trial) among South African adolescents and their primary caregivers. *Child Abuse & Neglect, 82*, 45-58. doi:10.1016/j.chiabu.2018.05.022

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS medicine, 6*(7), e1000097. doi:10.1371/journal.pmed.1000097

Mokkink, L. B., de Vet, H. C. W., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2018). COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Quality of Life Research, 27*(5), 1171-1179. doi:10.1007/s11136-017-1765-4

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., . . . de Vet, H. C. W. (2010a). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of Life Research, 19*(4), 539-549. doi:10.1007/s11136-010-9606-8

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., . . . de Vet, H. C. W. (2010b). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology, 63*(7), 737-745. doi:10.1016/j.jclinepi.2010.02.006

Negriff, S., Schneiderman, J. U., & Trickett, P. K. (2017). Concordance Between Self-Reported Childhood Maltreatment Versus Case Record Reviews for Child Welfare–Affiliated Adolescents: Prevalence Rates and Associations With Outcomes. *Child Maltreatment, 22*(1), 34-44. doi:10.1177/1077559516674596

Patrick, D. L., Burke, L. B., Gwaltney, C. J., Leidy, N. K., Martin, M. L., Molsen, E., & Ring, L. (2011). Content Validity—Establishing and Reporting the Evidence in Newly Developed Patient-Reported Outcomes (PRO) Instruments for Medical Product Evaluation: ISPOR PRO Good Research Practices Task Force Report: Part 2—Assessing Respondent Understanding. *Value in Health, 14*(8), 978-988. doi:10.1016/j.jval.2011.06.01

Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research, 27*(5), 1147-1157. doi:10.1007/s11136-018-1798-3

Prinsen, C. A. C., Vohra, S., Rose, M. R., Boers, M., Tugwell, P., Clarke, M., . . . Terwee, C. B. (2016). How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" – a practical guideline. *Trials, 17*(1), 449. doi:10.1186/s13063-016-1555-2

Ricci, L., Lanfranchi, J., Lemetayer, F., Rotonda, C., Guillemin, F., Coste, J., & Spitz, E. (2018). Qualitative Methods Used to Generate Questionnaire Items: A Systematic Review. *Qualitative Health Research, 29*(1), 149-156. doi:10.1177/1049732318783186

Rodriguez, C. M., Russa, M. B., & Harmon, N. (2011). Assessing abuse risk beyond self-report: Analog task of acceptability of parent-child aggression. *Child Abuse & Neglect, 35*(3), 199-209. doi:10.1016/j.chiabu.2010.12.004

Runyan, D. K., Dunne, M. P., Zolotor, A. J., Madrid, B., Jain, D., Gerbaka, B., . . . Choo, W. Y. (2009). The development and piloting of the ISPCAN Child Abuse Screening Tool—Parent version (ICAST-P). *Child Abuse & Neglect, 33*(11), 826-832. doi: 10.1016/j.chiabu.2009.09.006

Russa, M. B., & Rodriguez, C. M. (2010). Physical discipline, escalation, and child abuse potential: psychometric evidence for the Analog Parenting Task. *Aggressive Behavior, 36*(4), 251-260. doi:10.1002/ab.20345

Russell, B. S. (2010). Revisiting the Measurement of Shaken Baby Syndrome Awareness. *Child Abuse & Neglect, 34*(9), 671-676. doi:10.1016/j.chiabu.2010.02.008

Russell, B. S., & Britner, P. A. (2006). Measuring Shaken Baby Syndrome Awareness: Preliminary Reliability of a Caregiver Attitudes and Beliefs Survey. *Journal of Child and Family Studies, 15*(6), 765-777. doi:10.1007/s10826-006-9050-0

Saini, S. M., Hoffmann, C. R., Pantelis, C., Everall, I. P., & Bousman, C. A. (2019). Systematic review and critical appraisal of child abuse measurement instruments. *Psychiatry Research, 272*, 106-113. doi:10.1016/j.psychres.2018.12.068

Scholtes, V. A., Terwee, C. B., & Poolman, R. W. (2011). What makes a measurement instrument valid and reliable? *Injury, 42*(3), 236-240. doi:10.1016/j.injury.2010.11.042

Sedlak, A. J., Mettenburg, J., Basena, M., Petta, I., McPherson, K., Greene, A., & Li, S. (2010). *Fourth National Incidence Study of Child Abuse and Neglect (NIS–4): Report to Congress*. Washington, DC: Administration for Children and Families.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's Alpha. *Psychometrika, 74*(1), 107-120. doi:10.1007/s11336-008-9101-0

Speyer, R., Cordier, R., Kertscher, B., & Heijnen, B. J. (2014). Psychometric properties of questionnaires on functional health status in oropharyngeal dysphagia: a systematic literature review. *BioMed Research International, 2014*, 458678. doi:10.1155/2014/458678

Sprangers, M. A. G., & Aaronson, N. K. (1992). The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease: A review. *Journal of Clinical Epidemiology, 45*(7), 743-760. doi:10.1016/0895-4356(92)90052-O

Stewart, C., Kirisci, L., Long, A. L., & Giancola, P. R. (2015). Development and Psychometric Evaluation of the Child Neglect Questionnaire. *Journal of interpersonal violence, 30*(19), 3343-3366. doi:10.1177/0886260514563836

Stith, S. M., Liu, T., Davies, L. C., Boykin, E. L., Alder, M. C., Harris, J. M., . . . Dees, J. E. M. E. G. (2009). Risk factors in child maltreatment: A meta-analytic review of the literature. *Aggression and Violent Behavior, 14*(1), 13-29. doi:10.1016/j.avb.2006.03.006

Stoltenborgh, M., Bakermans-Kranenburg, M. J., Alink, L. R. A., & Ijzendoorn, M. H. (2015). The Prevalence of Child Maltreatment across the Globe: Review of a Series of Meta-Analyses. *Child Abuse Review, 24*(1), 37-50. doi:10.1002/car.2353

Straus, M. A., Hamby, S. L., Finkelhor, D., Moore, D. W., & Runyan, D. (1998). Identification of Child Maltreatment With the Parent-Child Conflict Tactics Scales: Development and Psychometric Data for a National Sample of American Parents. *Child Abuse & Neglect, 22*(4), 249-270. doi:10.1016/S0145-2134(97)00174-9

Straus, M. A., Hamby, S. L., & Warren, W. L. (2003). *The Conflict Tactics Scales Handbook: Revised Conflict Tactics Scales (CTS2) and CTS—Parent-child version (CTSPC)*. Los Angeles: Western Psychological Services.

Straus, M. A., Kinard, E. M., & Williams, L. M. (1995). *The multidimensional neglectful behavior scale, Form A: Adolescent and adult-recall version*. Durham, NH: Family Research Laboratory, University of New Hampshire.

Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health measurement scales : a practical guide to their development and use* (5th ed.). Oxford, UK: Oxford University Press.

Teicher, M. H., Samson, J. A., Anderson, C. M., & Ohashi, K. (2016). The effects of childhood maltreatment on brain structure, function and connectivity. *Nature Reviews Neuroscience, 17*(10), 652-666. doi:10.1038/nrn.2016.111

Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., . . . de Vet, H. C. W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology, 60*(1), 34-42. doi:doi.org/10.1016/j.jclinepi.2006.03.012

Terwee, C. B., Prinsen, C. A. C., Chiarotto, A., de Vet, H. C. W., Bouter, L. M., Alonso, J., . . . Mokkink, L. B. (2018). *COSMIN methodology for assessing the content validity of PROMs-User manual (version 1.0)*. Retrieved from https://www.cosmin.nl/wp-content/uploads/COSMIN-methodology-for-content-validity-user-manual-v1.pdf

Terwee, C. B., Prinsen, C. A. C., Chiarotto, A., Westerman, M. J., Patrick, D. L., Alonso, J., . . . Mokkink, L. B. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Quality of Life Research, 27*(5), 1159-1170. doi:10.1007/s11136-018-1829-0

Twentyman, C. T., Plotkin, R., Dodge, D., & Rohrbeck, C. A. (1981, November). *Inappropriate expectations of parents who maltreat their children*. Paper presented at the Annual Meeting of the Associationfor Advancement of Behavior Therapy, Toronto.

van Harmelen, A., van Tol, M., van der Wee, N. J. A., Veltman, D. J., Aleman, A., Spinhoven, P., . . . Elzinga, B. M. (2010). Reduced Medial Prefrontal Cortex Volume in Adults Reporting Childhood Emotional Maltreatment. *Biological Psychiatry, 68*(9), 832-838. doi:10.1016/j.biopsych.2010.06.011

Vittrup, B., Holden, G. W., & Buck, J. (2006). Attitudes predict the use of physical punishment: a prospective study of the emergence of disciplinary practices. *Pediatrics, 117*(6), 2055-2064. doi:10.1542/peds.2005-2204

Voisine, S., & Baker, A. J. L. (2012). Do universal parenting programs discourage parents from using corporal punishment: a program review. *Families in Society: The Journal of Contemporary Social Services, 93*(3), 212-218. doi:10.1606/1044-3894.4217

Wiering, B., de Boer, D., & Delnoij, D. (2017). Patient involvement in the development of patient-reported outcome measures: a scoping review. *Health Expect, 20*(1), 11-23. doi:10.1111/hex.12442

World Health Organization. (1999). *Report of the Consultation on Child Abuse Prevention*. Geneva, Switzerland: Author. Retrieved from https://apps.who.int/iris/handle/10665/65900

World Health Organization. (2016). *INSPIRE: Seven Strategies for Ending Violence Against Children*. Geneva, Switzerland: Author. Retrieved from http://apps.who.int/iris/bitstream/10665/207717/1/9789241565356-eng.pdf?ua=1

Zaidi, L. Y., Knutson, J. F., & Mehm, J. G. (1989). Transgenerational Patterns of Abusive Parenting - Analog and Clinical-Tests. *Aggressive Behavior, 15*(2), 137-152. doi:10.1002/1098-2337(1989)15:2<137::AID-AB2480150202>3.0.CO;2-O

Tables

Table 1. *Characteristics of the Included Instruments for the Assessment of Child Maltreatment.*

Instrument *(Acronym)*	Main constructs (Subscales)	Target population (child age)	Purpose of use	Number of subscales (total number of items); Range of score	Response Options	Recall period
Adult Adolescent Parenting Inventory-2 (AAPI-2) (Bavolek & Keene, 1999; Bavolek, Kline, McLaughlin, & Publicover, 1979)	Abusive and neglecting parenting practices (Inappropriate parental expectations; Parental lack of an empathic awareness of children's needs; Strong belief in the use and value of corporal punishment; Parent child role reversal; Oppressing children's power and independence)	Current and prospective parent populations (NR)	To provide prevalence estimates of child maltreatment; to screen child maltreatment; to evaluate prevention and treatment of physical and psychological child abuse	5 (40); Range: 0-50 (Raw total scores per subscale are converted into standard scores: range 0-10)	5-point ordinal scale (strongly disagree = 1 to strongly disagree = 5)	Not specified
Analog Parenting Task (APT) (Russa & Rodriguez, 2010; Zaidi, Knutson, & Mehm, 1989)	Attitude towards physical discipline (Physical discipline score: frequency of physical disciplinary response to alter children's behavior; Escalation score: frequency of switching from nonphysical to physical disciplinary tactics when child persisting in behavior)	Prospective parent populations (NR)	To identify high-risk pre-parent populations for primary prevention programming	2 (26); Range: 0-26	10 nominal scale (from nonphysical discipline tactics to physical discipline tactics)	Not specified
Child Neglect Questionnaire (CNQ) (Stewart, Kirisci, Long, & Giancola, 2015)	Child neglect (Physical neglect; Emotional neglect; Educational neglect; Supervision neglect)	Parents with older children (ages 10-12)	To detect children at high risk for parental neglect	4 (46); Range: 46-184	4-point ordinal scale (always = 1 to never = 4)	Past six months
Child Neglect Scales-Maternal Monitoring and Supervision scale (CNS-MMS) (Kirisci, Dunn, Mezzich, & Tarter, 2001; Loeber, Farrington, Stouthamer-Loeber, & Van Kammen, 1998)	Child neglect by parents	Mothers (NR)	To quantify severity of child neglect by mothers	1 (11); Range: 11-33	3-point ordinal scale (hardly ever = 1 to often = 3)	Past six months
Child Trauma Screen-Exposure Score (CTS-ES) (Lang & Connell, 2017)	Potentially traumatic event, including childhood physical abuse, sexual abuse, and domestic or community violence	Caregivers with children (ages over 6)	To screen children for trauma exposure	1 (4); Range: 0-4	Dichotomous scale (No = 0 or Yes = 1)	Not specified

(continued)

Table 1. *(continued)*

Instrument *(Acronym)*	Main constructs (Subscales)	Target population (child age)	Purpose of use	Number of subscales (total number of items); Range of score	Response Options	Recall period
Conflict Tactics Scales: Parent-Child version (CTSPC) (Straus, Hamby, Finkelhor, Moore, & Runyan, 1998; Straus, Hamby, & Warren, 2003)	Physical and psychological child abuse (Nonviolent discipline; Psychological aggression; Physical assault) (Optional supplementary three subscales: Weekly discipline; Neglect; Sexual abuse)	Parents (NR)	To provide prevalence estimates of child maltreatment; to screen child maltreatment; to evaluate prevention and treatment of physical and psychological child abuse	3 (22); Range: 0-550 (Raw scores per item are converted into frequency scores: 0 = 0, 1 = 1, 2 = 2, 3–5 = 4, 6–10 = 8, 11–20 = 15, and > 20 = 25) (Supplementary subscales: 3 (13): 0-233)	8-point ordinal scale (0 = never happened; 1 = once in the past year; 2 = twice; 3 = 3–5 times; 4 = 6–10 times; 5 = 11–20 times; 6 = more than 20 times; 7 = not in the past year, but it happened before) (Supplementary subscales: 3 to 7-point ordinal scale)	Past one year (Optional supplementary subscales: past one week to lifetime before 18 years old)
Family Maltreatment-Child Abuse criteria (FM-CA) (Heyman, Snarr, Slep, Baucom, & Linkh, 2009)	Clinically significant child abuse and neglect (Physical child abuse; Psychological child abuse)	Parents (NR)	To screen clinically significant child abuse	2 (27); Range: 0-63	Dichotomous scale for physical child abuse subscale (I did = 0 or I never did = 1); 6-point ordinal scale for psychological child abuse subscale (never = 0 to more than once a day = 5)	Past one year
ISPCAN Child Abuse Screening Tool for use in Trials (ICAST-Trial) **(**Meinch et al., 2018; Runyan et al., 2009)	Child abuse and neglect (Physical abuse; Emotional abuse; Contact sexual abuse; Neglect)	Caregivers (ages 10-18)	To evaluate effectiveness of child abuse prevention program	4 (14) Range: 0-112	9-point ordinal scale (never = 0 to more than 8 times = 8)	Past one month
Intensity of Parental Punishment Scale (IPPS) (Gordon, Jones, & Nowicki, 1979)	Intensity of parent behavioral responses to hypothetical child misbehavior situations (School misbehavior; Disobedience after a recent reminder; Public disobedience; Crying; Destructiveness)	Parents of children (ages 5-10)	To provide investigators with cost effective information of long-term effects on parental punishments than time consuming interview and observation without any demonstrable reduction in accuracy	5 (33); Range: 33-231	7-point ordinal scale (no reaction = 1 to very strong punishment = 7)	Not specified
Mother-Child Neglect Scale (MCNS) (Lounds, Borkowski, & Whitman, 2004; Straus, Kinard, & Williams, 1995)	Maternal neglectful behavior towards their children (Emotional neglect; Cognitive neglect; Supervisory neglect; Physical needs neglect)	Mothers (NR)	To screen parents at highest risk for child neglect for prevention of its future occurrence	4 (20); Range: 20-80	4-point ordinal scale (strongly disagree = 1 to strongly agree = 4)	Past one year

(Continued)

Table 1. *(continued)*

Instrument *(Acronym)*	**Main constructs (Subscales)**	**Target population (child age)**	**Purpose of use**	**Number of subscales (total number of items); Range of score**	**Response Options**	**Recall period**
Mother-Child Neglect Scale-Short Form (MCNS-SF) (Lounds et al., 2004; Straus et al., 1995)	Maternal neglectful behavior towards their children (Emotional neglect; Cognitive neglect; Supervisory neglect; Physical needs neglect)	Mothers (NR)	To screen parents at highest risk for child neglect for prevention of its future occurrence	2 (8); Range: 4-32	4-point ordinal scale (strongly disagree = 1 to strongly agree = 4)	Past one year
Parent-Child Aggression Acceptability Movie Task (P-CAAM) (Rodriguez, Russa, & Harmon, 2011)	Acceptance of parent-child aggression (Physical discipline; Physical abuse)	Current and prospective parent populations (NR)	To assess intervention programming outcomes	2 (8 video clips: 90 sec each); Range: 0-NR	Clips builds towards 'initial physical contact between caregiver and child'; Rater should identify that moment and stop video; Delay between actual physical contact and stop video = score (per video)	Not specified
Parent Opinion Questionnaire (POQ) (Twentyman, Plotkin, Dodge, & Rohrbeck, 1981, November)	Parental expectations of child behavior (Self-care; Family responsibility and care of siblings; Help and affection to parents; Leaving children alone; Proper behavior and feelings; Punishment)	Parents (NR)	To identify abusive parents for child maltreatment service	6 (60); Range: 0-60	Dichotomous scale (disagree = 0 or agree = 1)	Not specified
Parental Response to Child Misbehavior questionnaire (PRCM) (Holden & Zambarano, 1992; Vittrup et al., 2006)	Discipline techniques used by parents in response to their children's misbehaviors.	Parents with young children (NR)	To obtain information regarding the frequency of specific discipline techniques	1 (12); Range: 0-72	6-point ordinal scale (never = 0 to 9 ≥ times per week = 6)	Past one week
Shaken Baby Syndrome awareness assessment-Short Version (SBS-SV) (Russell, 2010; Russell & Britner, 2006)	Shaken baby syndrome awareness (Soothing techniques; Discipline techniques; Potential for injury)	Parents, babysitters, and childcare providers of young children (ages younger than 2)	To provide a measure for caregiver education and other service provision concerning the care of infants younger than two years	3 (36); Range: : 36-216	6-point ordinal scale (strongly disagree = 1 to strongly agree = 6)	Not specified

Notes. All information was derived from all eligible studies and the original included instruments; NR = Not Reported.

Table 2. *Methodological Quality Assessment of Development and Content Validity Studies on Content Validity of the Included Instruments.*

Instrument	Reference	Development study quality[a]			Content validity study quality[a]				
		Item generation[b]	Cognitive interview[b]		Asking parents/carers[b]			Asking professionals[b]	
		Relevance	Comprehensiveness	Comprehensibility	Relevance	Comprehensiveness	Comprehensibility	Relevance	Comprehensiveness
AAPI-2	Bavolek et al. (1979)	NR	Inadequate (4.8%)	Inadequate (21.6%)	NR	NR	Doubtful (42.9%)	Doubtful (40.0%)	NR
APT	Zaidi et al. (1989)	NR	NR	NR	NR	NR	NR	NR	NR
CNQ	Stewart et al. (2015)	NR	NR	NR	NR	NR	NR	Doubtful (33.3%)	NR
CNS-MMS	Loeber et al. (1998)	NR	NR	NR	NR	NR	NR	NR	NR
CTS-ES	Lang and Connell (2017)	NR	NR	NR	NR	NR	NR	Doubtful (33.3%)	NR
CTSPC	Straus et al. (1998)	Inadequate (20.0%)	Inadequate (7.1%)	Doubtful (36.4%)	NR	NR	Doubtful (33.3%)	NR	NR
FM-CA	Heyman et al. (2019)	Doubtful (50.0%)	Inadequate (9.5%)	Inadequate (9.5%)	Doubtful (38.1%)	NR	Adequate (66.6%)	NR	NR
ICAST-Trial	Runyan et al. (2009)	NR	NR	NR	NR	NR	NR	NR	NR
	Meinch et al. (2018)	NR	NR	NR	Very good (76.2%)	NR	Very good (76.2%)	NR	NR
IPPS	Gordon et al. (1979)	Inadequate (3.5%)	Inadequate (7.1%)	Inadequate (4.8%)	Inadequate (12.5%)	NR	NR	Doubtful (33.3%)	NR
MCNS	Straus et al. (1995)	NR	NR	NR	NR	NR	NR	NR	NR
MCNS-SF	Straus et al. (1995)	NR	NR	NR	NR	NR	NR	NR	NR
P-CAAM	Rodriguez et al. (2011)	NR	NR	NR	NR	NR	NR	Doubtful (40.0%)	NR
POQ	Twentyman et al. (1981, November)	NR	NR	NR	Doubtful (38.1%)	NR	NR	Doubtful (40.0%)	NR
PRCM	Holden and Zambarano (1992)	NR	NR	NR	NR	NR	NR	NR	NR
SBS-SV	Russell and Britner (2006)	NR	Inadequate (7.1%)	Inadequate (7.1%)	NR	NR	NR	Doubtful (33.3%)	NR

Notes. AAPI-2 = Adult Adolescent Parenting Inventory-2; APT = Analog Parenting Task; CNQ = Child Neglect Questionnaire; CNS-MMS = Child Neglect Scales-Maternal Monitoring and Supervision scale; CTS-ES = Child Trauma Screen-Exposure Score; CTSPC = Conflict Tactics Scales: Parent-Child version; FM-CA = Family Maltreatment-Child Abuse criteria; ICAST-Trial = ISPCAN (International Society for the Prevention of Child Abuse and Neglect) Child Abuse Screening Tool for use in Trials; IPPS = Intensity of Parental Punishment Scale; MCNS = Mother-Child Neglect Scale; MCNS-SF = Mother-Child Neglect Scale-Short Form; P-CAAM = Parent-Child Aggression Acceptability Movie task; POQ = Parent Opinion Questionnaire; PRCM = Parental Response to Child Misbehavior questionnaire; SBS-SV = Shaken Baby Syndrome awareness assessment-Short Version.

[a] The methodological quality per development and content validity study was rated using the COSMIN checklist (Mokkink et al., 2010a). The overall methodological quality per study was presented as a percentage of the ratings (Cordier et al., 2015): Inadequate = 0-25%; Doubtful = 25.1-50%; Adequate = 50.1-75%; Very good = 75.1-100%; NR = Not Reported.

[b] The methodological quality was rated in the three aspects of content validity: relevance; comprehensiveness; and comprehensibility. The development study was rated in the two parts (item generation and cognitive interview); the content validity study was rated in the two study categories asking parents/carers or experts about the relevance, comprehensiveness, and comprehensibility.

Table 3. *Quality of Content Validity per Development and Content Validity Study, and Content of Instrument Itself.*

Instrument	Reference	Relevance[a]			Comprehensiveness[a]			Comprehensibility[a]		
		Development study	Content validity study	Content of instrument	Development study	Content validity study	Content of instrument	Development study	Content validity study	Content of instrument
AAPI-2	Bavolek et al. (1979)	?	?	+	?	?	+	?	?	+
APT	Zaidi et al. (1989)	?	?	+	?	?	+	?	?	+
CNQ	Stewart et al. (2015)	?	?	+	?	?	+	?	?	+
CNS-MMS	Loeber et al. (1998)	?	?	+	?	?	+	?	?	+
CTS-ES	Lang and Connell (2017)	?	?	±	?	?	-	?	?	+
CTSPC	Straus et al. (1998)	?	?	+	?	?	+	?	?	+
FM-CA	Heyman et al. (2019)	+	?	+	?	?	+	?	+	+
ICAST-Trial	Meinch et al. (2018); Runyan et al. (2009)	?	?	+	?	?	+	?	?	+
IPPS	Gordon et al. (1979)	?	?	+	?	?	+	?	?	+
MCNS	Straus et al. (1995)	?	?	+	?	?	+	?	?	+
MCNS-SF	Straus et al. (1995)	?	?	+	?	?	+	?	?	+
P-CAAM	Rodriguez et al. (2011)	?	?	?	?	?	?	?	?	?
POQ	Twentyman et al. (1981, November)	?	?	±	?	?	-	?	?	+
PRCM	Holden and Zambarano (1992)	?	?	+	?	?	+	?	?	±
SBS-SV	Russell and Britner (2006)	?	?	+	?	?	+	?	?	+

Notes. AAPI-2 = Adult Adolescent Parenting Inventory-2; APT = Analog Parenting Task; CNQ = Child Neglect Questionnaire; CNS-MMS = Child Neglect Scales-Maternal Monitoring and Supervision scale; CTS-ES = Child Trauma Screen-Exposure Score; CTSPC = Conflict Tactics Scales: Parent-Child version; FM-CA = Family Maltreatment-Child Abuse criteria; ICAST-Trial = ISPCAN (International Society for the Prevention of Child Abuse and Neglect) Child Abuse Screening Tool for use in Trials; IPPS = Intensity of Parental Punishment Scale; MCNS = Mother-Child Neglect Scale; MCNS-SF = Mother-Child Neglect Scale-Short Form; P-CAAM = Parent-Child Aggression Acceptability Movie Task; POQ = Parent Opinion Questionnaire; PRCM = Parental Response to Child Misbehavior questionnaire; SBS-SV = Shaken Baby Syndrome awareness assessment-Short Version.

[a] The quality of content validity (relevance, comprehensiveness, and comprehensibility) per study and content of instrument was rated using the criteria for good content validity (Terwee, Prinsen, Chiarotto, de Vet, et al., 2018): + = Sufficient rating; ? = Indeterminate rating; - = Insufficient rating; ± = Inconsistent rating. Rating for development and content validity studies was determined based on the data from development and content validity studies; rating for content of instrument was determined based on reviewers' subjective opinion on content of instrument itself (items and instructions).

Table 4. *Overall Quality of Content Validity and Evidence Quality per Instrument.*

Instrument	Relevance		Comprehensiveness		Comprehensibility	
	Overall quality of content validity[a]	Quality of evidence[b]	Overall quality of content validity[a]	Quality of evidence[b]	Overall quality of content validity[a]	Quality of evidence[b]
AAPI-2	+	Moderate	+	Very low	+	Very low
APT	+	Very low	+	Very low	+	Very low
CNQ	+	Moderate	+	Very low	+	Very low
CNS-MMS	+	Very low	+	Very low	+	Very low
CTS-ES	±	Low	-	Very low	+	Very low
CTSPC	+	Very low	+	Low	+	Low
FM-CA	+	Moderate	+	Very low	+	High
ICAST-Trial	+	High	+	Very low	+	High
IPPS	+	Moderate	+	Very low	+	Very low
MCNS	+	Very low	+	Very low	+	Very low
MCNS-SF	+	Very low	+	Very low	+	Very low
P-CAAM	?	NE	?	NE	?	NE
POQ	±	Low	-	Very low	+	Very low
PRCM	+	Very low	+	Very low	±	Very low
SBS-SV	+	Low	+	Very low	+	Very low

Notes. AAPI-2 = Adult Adolescent Parenting Inventory-2; APT = Analog Parenting Task; CNQ = Child Neglect Questionnaire; CNS-MMS = Child Neglect Scales-Maternal Monitoring and Supervision scale; CTS-ES = Child Trauma Screen-Exposure Score; CTSPC = Conflict Tactics Scales: Parent-Child version; FM-CA = Family Maltreatment-Child Abuse criteria; ICAST-Trial = ISPCAN (International Society for the Prevention of Child Abuse and Neglect) Child Abuse Screening Tool for use in Trials; IPPS = Intensity of Parental Punishment Scale; MCNS = Mother-Child Neglect Scale; MCNS-SF = Mother-Child Neglect Scale-Short Form; P-CAAM = Parent-Child Aggression Acceptability Movie Task; POQ = Parent Opinion Questionnaire; PRCM = Parental Response to Child Misbehavior questionnaire; SBS-SV = Shaken Baby Syndrome awareness assessment-Short Version.

[a] The overall quality of content validity (relevance, comprehensiveness, and comprehensibility) was determined by qualitatively summarizing all ratings on content validity per study of each instrument and reviewers' ratings on content of instrument itself (Terwee, Prinsen, Chiarotto, de Vet, et al., 2018): + = Sufficient rating; ? = Indeterminate rating; - = Insufficient rating; ± = Inconsistent rating.

[b] The quality of evidence (confidence level for the overall quality rating of content validity) was rated using a modified GRADE approach (Terwee, Prinsen, Chiarotto, Westerman, et al., 2018); High = high level of confidence; Moderate = moderate level of confidence; Low = low level of confidence; Very Low = very low level of confidence; NE = Not Evaluated (instruments could not be retrieved).

Figures

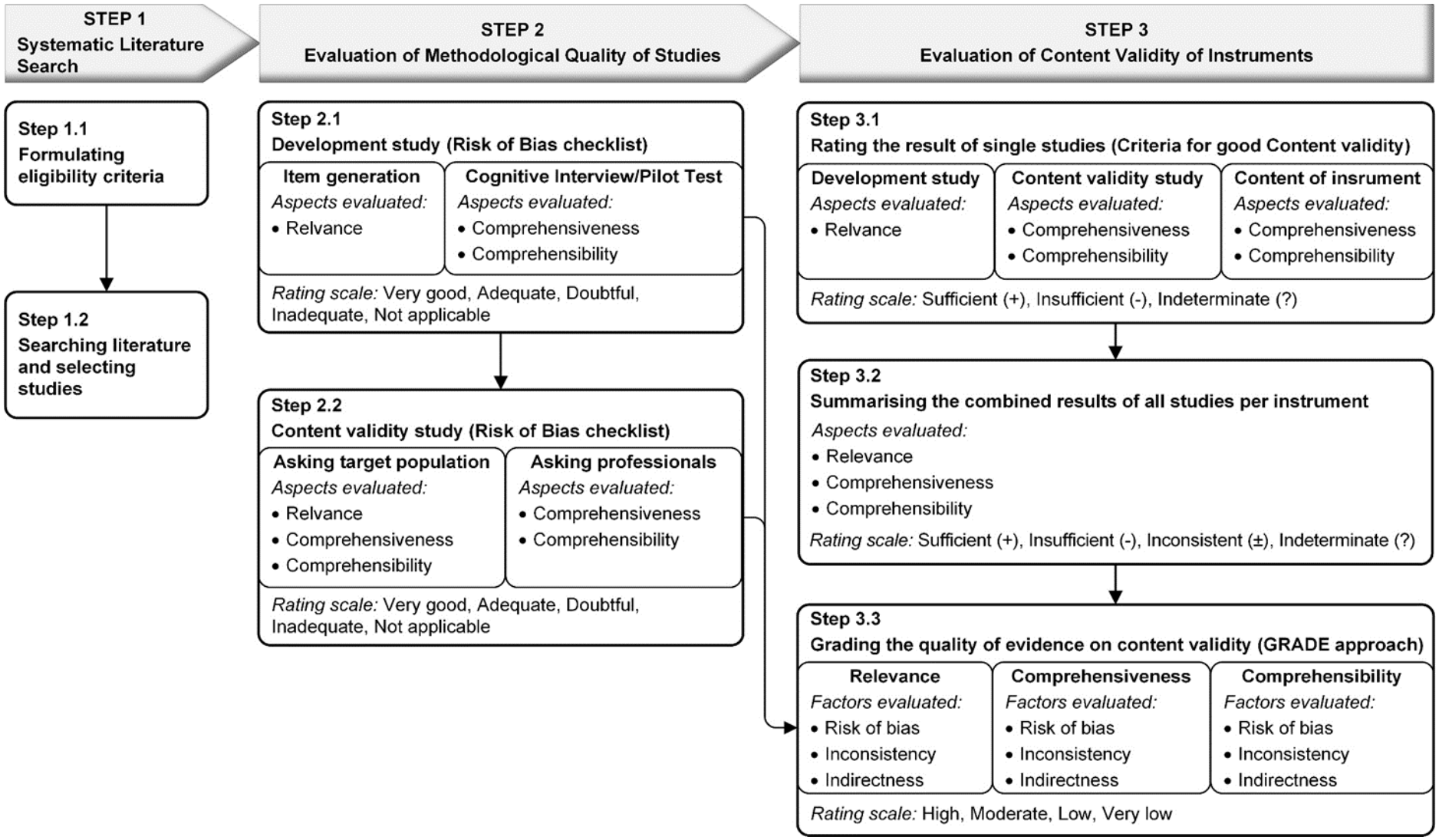
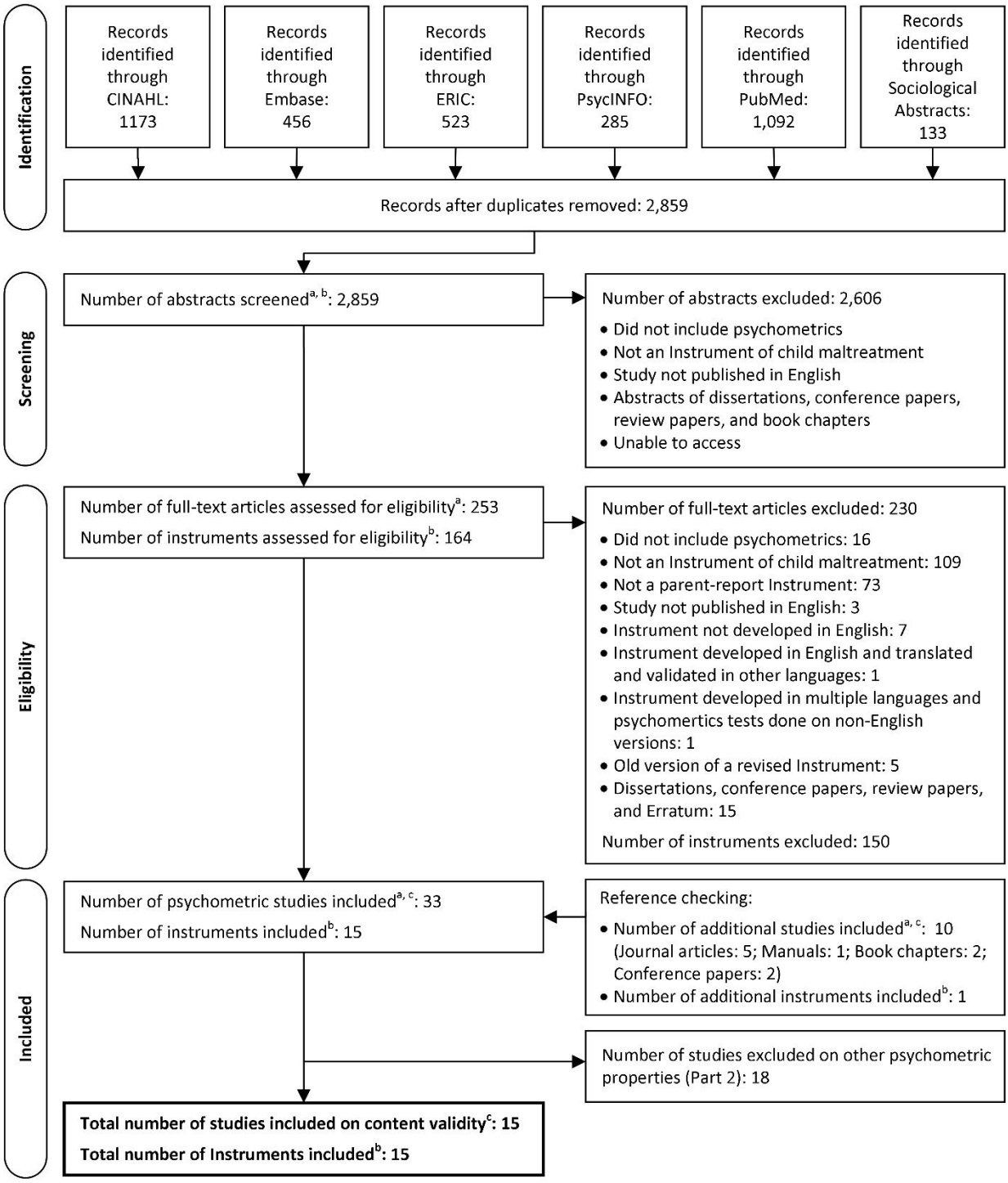


Figure 1. Study design: steps for PRISMA and COSMIN processes.



Notes. The literature searches and study selection were conducted for both this paper on content validity (Part 1) and a companion paper on other psychometric properties (Part 2).

[a] Studies on any psychometric property were eligible if they: (1) were journal articles and manuals published in English; (2) reported on psychometric data of any psychometric properties of eligible instruments.

[b] Instruments were eligible if: (1) attitude towards child maltreatment or maltreating behaviors towards children was assessed; (2) parents or caregivers were asked on their own attitude or behaviors; (3) at least one subscale or a minimum of 30% of all items referred to one or more types of child maltreatment (physical abuse, emotional abuse, neglect, and sexual abuse); and (4) instruments were developed and published in English.

[c] Studies on content validity were eligible if they: (1) were any type of study (e.g., journal articles, book chapters, conference papers) published in English; (2) reported on the development of new items of eligible instruments; and/or (3) evaluated the relevance, comprehensiveness, or comprehensibility of the content of the eligible instruments by asking parents/caregivers, or professionals.

Figure 2. Flow diagram of the reviewing procedure based on PRISMA (Moher et al., 2009).

Appendices

Appendix A. *Database Search Strategies.*

Database	Search Terms (Subject heading and Free text words)	Number of records
CINAHL	(((MH "Child Abuse+") OR (MH "Domestic Violence+") OR (MH "Family Conflict") OR (MH "Aggression+") OR (MH "Punishment")) AND ((MH "Parents+") OR (MH "Parenting") OR (MH "Father-Infant Relations")OR (MH "Father-Child Relations") OR (MH "Fathers+") OR (MH "Mother-Child Relations") OR (MH "Mother-Infant Relations") OR (MH "Mothers+") OR (MH "Family+") OR (MH "Caregivers") OR (MH "Child Rearing+")) AND ((MH "Psychometrics") OR (MH "Measurement Issues and Assessments") OR (MH "Validity") OR (MH "Predictive Validity") OR (MH "Reliability and Validity") OR (MH "Internal Validity") OR (MH "Face Validity") OR (MH "External Validity") OR (MH "Discriminant Validity") OR (MH "Criterion-Related Validity") OR (MH "Consensual Validity") OR (MH "Concurrent Validity") OR (MH "Qualitative Validity") OR (MH "Construct Validity") OR (MH "Content Validity") OR (MH "Questionnaire Validation") OR (MH "Validation Studies") OR (MH "Test-Retest Reliability") OR (MH "Sensitivity and Specificity") OR (MH "Reproducibility of Results") OR (MH "Reliability") OR (MH "Intrarater Reliability") OR (MH "Interrater Reliability") OR (MH "Measurement Error") OR (MH "Bias (Research)") OR (MH "Selection Bias") OR (MH "Sampling Bias") OR (MH "Precision") OR (MH "Sample Size Determination") OR (MH "Repeated Measures") OR (Psychometric* or reliability or validit* or reproducibility or bias))) OR (((child OR children OR infant* OR toddler* OR neonate* OR baby OR babies OR adolescent* OR teen* OR minor*) AND (victim* OR aggress* OR punish* OR abus* OR maltreat* OR neglect* OR mistreat* or violen* or conflict* or batter* or molest*) AND (rear* OR parent* OR father* OR mother* OR family OR families OR domestic* OR caregiver* OR carer* OR caring OR home OR homes) AND (psychometric* OR reliabilit* OR validit* OR reproducibilit* OR bias)) Limiters - Published Date: 20181001-20191031)	1173
Embase	((child abuse/ OR child neglect/ OR emotional abuse/ OR physical abuse/ OR battering/ OR domestic violence/ OR physical violence/ OR family conflict/ OR victim/ OR aggression/ OR punishment/) AND (parent/ OR father/ OR father child relation/ OR mother/ OR mother child relation/ OR family/ OR caregiver/ OR child rearing/) AND (psychometry/ or validity/ or reliability/ or measurement error/ or measurement precision/ or measurement repeatability/ or error/ or statistical bias/ or test retest reliability/ or intrarater reliability/ or interrater reliability/ or accuracy/ or criterion validity/ or internal validity/ or face validity/ or external validity/ or discriminant validity/ or concurrent validity/ or qualitative validity/ or construct validity/ or content validity/)) OR (((child OR children OR infant* OR toddler* OR neonate* OR baby OR babies OR adolescent* OR teen* OR minor*) AND (victim* OR aggress* OR punish* OR abus* OR maltreat* OR neglect* OR mistreat* or violen* or conflict* or batter* or molest*) AND (rear* OR parent* OR father* OR mother* OR family OR families OR domestic* OR caregiver* OR carer* OR caring OR home OR homes) AND (psychometric* OR reliabilit* OR validit* OR reproducibilit* OR bias)) limit to yr="2019 -Current")	456
ERIC	((Child abuse/ OR Child neglect/ OR violence/ OR family violence/) AND (parenting styles/ OR parents/ OR child rearing/ OR father attitudes/ OR fathers/ OR mother attitudes/ OR mothers/ OR family attitudes/ OR caregiver attitudes/ OR caregiver child relationship/ OR caregiver role/ OR family environment/) AND (Psychometrics/ OR Validity/ OR Reliability/ OR Error of Measurement/ OR Bias/ OR Interrater Reliability/ OR Accuracy/ OR Predictive Validity/ OR Construct Validity/ OR Content Validity/)) OR (((child OR children OR infant* OR toddler* OR neonate* OR baby OR babies OR adolescent* OR teen* OR minor*) AND (victim* OR aggress* OR punish* OR abus* OR maltreat* OR neglect* OR mistreat* or violen* or conflict* or batter* or molest*) AND (rear* OR parent* OR father* OR mother* OR family OR families OR domestic* OR caregiver* OR carer* OR caring OR home OR homes) AND (psychometric* OR reliabilit* OR validit* OR reproducibilit* OR bias)) limit to yr="Last year")	523

(continued)

Appendix A. *(continued)*

Database	Search Terms (Subject heading and Free text words)	Number of records
PsycINFO	((child abuse/ OR child neglect/ OR violence/ OR domestic violence/ OR physical abuse/ OR family conflict/ OR victimization/ OR aggressive behaviOR/ OR aggressiveness/ OR punishment/) AND (parent child communication/ OR parent child relations/ OR parenting/ OR parenting style/ OR parents/ OR father child communication/ OR father child relations/ OR fathers/ OR mother child communication/ OR mother child relations/ OR mothers/ OR family/ OR caregivers/) AND (Psychometrics/ OR Statistical Validity/ OR Test Validity/ OR Statistical Reliability/ OR Test Reliability/ OR Error of Measurement/ OR Errors/ OR Response Bias/ OR Interrater Reliability/ OR Repeated Measures/)) OR (((child OR children OR infant* OR toddler* OR neonate* OR baby OR babies OR adolescent* OR teen* OR minor*) AND (victim* OR aggress* OR punish* OR abus* OR maltreat* OR neglect* OR mistreat* or violen* or conflict* or batter* or molest*) AND (rear* OR parent* OR father* OR mother* OR family OR families OR domestic* OR caregiver* OR carer* OR caring OR home OR homes) AND (psychometric* OR reliabilit* OR validit* OR reproducibilit* OR bias)) limit to yr= “2019 -Current”)	285
PubMed	(("Child Abuse"[Mesh] OR "Physical Abuse"[Mesh] OR "Domestic Violence"[Mesh] OR "Violence"[Mesh] OR "Family Conflict"[Mesh] OR "Aggression"[Mesh] OR "Punishment"[Mesh]) AND ("Parents"[Mesh] OR "Parent-Child Relations"[Mesh] OR "Parenting"[Mesh] OR "Fathers"[Mesh] OR "Father-Child Relations"[Mesh] OR "Mothers"[Mesh] OR "Mother-Child Relations"[Mesh] OR "Family"[Mesh] OR "Caregivers"[Mesh] OR "Child Rearing"[Mesh]) AND ("Psychometrics"[Mesh] OR "Reproducibility of Results"[Mesh] OR "Validation Studies as Topic"[Mesh] OR "Validation Studies" [Publication Type] OR "Bias"[Mesh] OR "Observer Variation"[Mesh] OR "Selection Bias"[Mesh] OR "Diagnostic Errors"[Mesh] OR "Dimensional Measurement Accuracy"[Mesh] OR “Predictive Value of Tests"[Mesh] OR "Discriminant Analysis"[Mesh])) OR (((child OR children OR infant* OR toddler* OR neonate* OR baby OR babies OR adolescent* OR teen* OR minor*) AND (victim* OR aggress* OR punish* OR abus* OR maltreat* OR neglect* OR mistreat* or violen* or conflict* or batter* or molest*) AND (rear* OR parent* OR father* OR mother* OR family OR families OR domestic* OR caregiver* OR carer* OR caring OR home OR homes) AND (psychometric* OR reliabilit* OR validit* OR reproducibilit* OR bias)) Filters: Publication date from 2018/10/05 to 2019/10/05)	1092
Sociological Abstracts	(MAINSUBJECT.EXACT("Child Neglect") OR MAINSUBJECT.EXACT("Child Abuse") OR (MAINSUBJECT.EXACT("Violence") OR MAINSUBJECT.EXACT("Family Violence")) OR MAINSUBJECT.EXACT("Family Conflict") OR MAINSUBJECT.EXACT("Victimization") OR MAINSUBJECT.EXACT("Victims") OR MAINSUBJECT.EXACT("Aggression") OR (MAINSUBJECT.EXACT("Punishment") OR MAINSUBJECT.EXACT("Corporal Punishment")) OR MAINSUBJECT.EXACT("Emotional Abuse")) AND (MAINSUBJECT.EXACT("Parent Child Relations") OR MAINSUBJECT.EXACT("Parental Influence") OR MAINSUBJECT.EXACT("Parents") OR MAINSUBJECT.EXACT("Parental Attitudes") OR MAINSUBJECT.EXACT("Parenthood")) OR MAINSUBJECT.EXACT("Childrearing Practices") OR MAINSUBJECT.EXACT("Fathers") OR MAINSUBJECT.EXACT("Mothers") OR (MAINSUBJECT.EXACT("Family") OR MAINSUBJECT.EXACT("Family Relations") OR MAINSUBJECT.EXACT("Family Conflict") OR MAINSUBJECT.EXACT("Family Violence")) OR MAINSUBJECT.EXACT("Caregivers")) AND (MAINSUBJECT.EXACT("Psychometric Analysis") OR MAINSUBJECT.EXACT("Validity") OR MAINSUBJECT.EXACT("Reliability") OR MAINSUBJECT.EXACT("Error of Measurement") OR MAINSUBJECT.EXACT("Errors") OR MAINSUBJECT.EXACT("Test Bias") OR MAINSUBJECT.EXACT("Statistical Bias") OR MAINSUBJECT.EXACT("Bias") OR MAINSUBJECT.EXACT("Accuracy") OR MAINSUBJECT.EXACT("Agreement") OR MAINSUBJECT.EXACT("Research Design Error") OR MAINSUBJECT.EXACT("Specificity") OR MAINSUBJECT.EXACT("Sampling"))	133

Notes. All searches performed on the 29th of January 2018 with an update on the 5th of October 2019.

Appendix B. *Overview of Child Maltreatment Instrument: Reasons for Exclusion.*

No	Instrument[a] *(alphabetical order)*	Abbreviation	Reason for exclusion
1	Adolescent Clinical Sexual Behavior Inventory (William N. Friedrich, Lysne, Sim, & Shamos, 2004)	ACSBI	Not a measure of child maltreatment
2	Adolescent Sexual Behavior Inventory- Self Report (Wherry, Berres, Sim, & Friedrich, 2009)	ACSBI-S	Not a measure of child maltreatment
3	Adult Attachment Interviews (Hesse, 2008)	AAIs	Not a parent-report measure
4	Adult-Adolescent Parenting Inventory (Bavolek, 1984)	AAPI	Old version of a revised measure
5	Adverse Childhood Experiences Questionnaire (Felitti et al., 1998)	ACEs	Not a parent-report measure
6	Alabama Parenting Questionnaire (Shelton, Frick, & Wootton, 1996)	APQ	Not a measure of child maltreatment
7	Assessing Environments (Berger, Knutson, Mehm, & Perkins, 1988)	AEIII	Not a parent-report measure
8	Assessment of parental awareness of the shaken baby syndrome[b] (Mann, Rai, Sharif, & Vavasseur, 2015)	N/A	No psychometric data found
9	Body Image Victimization Experiences Scale (Duarte & Pinto-Gouveia, 2017)	BIVES	Not a measure of child maltreatment
10	Brief Child Abuse Potential Inventory (Ondersma, Chaffin, Mullins, & LeBreton, 2005)	BCAP	Not a measure of child maltreatment
11	Brigid Collins Risk Screener (Weberling, Forgays, Crain-Thoreson, & Hyman, 2003)	BCRS	Not a measure of child maltreatment
12	California Family Risk Assessment (W. L. Johnson, 2011)	CFRA	Not a parent-report measure
13	Caregiver–Child Social/Emotional and Relationship Rating Scale (McCall, Groark, & Fish, 2010)	CCSERRS	Not a measure of child maltreatment
14	Child Abuse Inventory at Emergency Rooms (Sittig et al., 2016)	CHAINER	Not a parent-report measure
15	Child Abuse Potential Inventory (Milner, 1986)	CAP	Not a measure of child maltreatment
16	Child Abuse Risk Assessment Scale (Chan, 2012)	CARAS	Not developed in English
17	Child and Adolescent Trauma Screen (Sachser et al., 2017)	CATS	Not a measure of child maltreatment
18	Child Behavior Checklist (Achenbach & Rescorla, 2000)	CBCL	Not a measure of child maltreatment
19	Child emotional maltreatment module[b] (A. M. Slep, Heyman, & Snarr, 2011)	N/A	No psychometric data found
20	Child maltreatment assessment (Salum et al., 2016)	N/A	Not developed in English
21	Child Maltreatment Measure[b] (Tajima, Herrenkohl, Huang, & Whitney, 2004)	N/A	No psychometric data found
22	Child Protective Services Review Document (Fanshel, Finch, & Grundy, 1994)	CPSRD	Not a parent-report measure
23	Child Reflective Functioning Scale (Ensink et al., 2015)	CRF	Not a measure of child maltreatment
24	Child Sexual Behavior Inventory (W. N. Friedrich et al., 2001)	CSBI	Not a measure of child maltreatment
25	Child Well-Being Scales (Gaudin, Polansky, & Kilpatrick, 1992)	CWBS	Not a parent-report measure
26	Childhood Experience of Care and Abuse (Brown, Craig, Harris, Handley, & Harvey, 2007)	CECA	Not a parent-report measure
27	Childhood Experience of Care and Abuse Questionnaire (N. Smith, Lam, Bifulco, & Checkley, 2002)	CECA.Q	Not a parent-report measure
28	Childhood Experiences of Violence Questionnaire (Walsh, MacMillan, Trocme, Jamieson, & Boyle, 2008)	CEVQ	Not a parent-report measure
29	Childhood Trauma Interview (Fink, Bernstein, Handelsman, Foote, & Lovejoy, 1995)	CTI	Not a parent-report measure
30	Childhood Trauma Questionnaire (Bernstein, Ahluvalia, Pogge, & Handelsman, 1997)	CTQ	Not a parent-report measure
31	Childhood Trauma Questionnaire Short Form (Forde, Baron, Scher, & Stein, 2012)	CTQ-SF	Not a parent-report measure
32	Child-Parent Relationship Scale (Driscoll & Pianta, 2011)	CPRS	Not a measure of child maltreatment
33	Child–Parent Relationship Scale–Short Form (Pianta, 1992)	CPRS-SF	Not a measure of child maltreatment
34	Children Intimate Relationships, and Conflictual Life Events Interview (Marshall, Feinberg, Jones, & Chote, 2017)	CIRCLE	Not a parent-report measure
35	Children's Impact of Traumatic Events Scale-Revised (Chaffin & Shultz, 2001)	CITES-R	Not a measure of child maltreatment
36	Christchurch Trauma Assessment (Nelson, Lynskey, Heath, & Martin, 2010)	N/A	Not a parent-report measure
37	Cleveland Child Abuse Potential Scale (Ezzo & Young, 2012)	C-CAPS	Not a parent-report measure
38	Comprehensive Childhood Maltreatment Inventory (Riddle & Aponte, 1999)	CCMI	Not a parent-report measure
39	Conflict Tactic Scale 2 (Straus et al., 2003)	CTS 2	Not a measure of child maltreatment

(Continued)

Appendix B. *(continued)*

No	Instrument[a] *(alphabetical order)*	Abbreviation	Reason for exclusion
40	Conflict Tactics Scales (Straus et al., 2003)	CTS	Not a measure of child maltreatment
41	Defense Style Questionnaire (Bond & Wesley, 1996)	DSQ	Not a parent-report measure
42	Disciplinary Methods Interview[b] (Thompson, 2017)	N/A	Not a measure of child maltreatment
43	Discipline Survey (Socolar, Savage, Devellis, & Evans, 2004)	N/A	Not a measure of child maltreatment
44	Dunedin Family Services Indicator (Muir et al., 1989)	DFSI	Not a parent-report measure
45	Dyadic Parent-Child Interaction Coding System-II (Eyberg, Bessmer, Newcomb, Edwards, & Robinson, 1994)	DPICS-II	Not a parent-report measure
46	Egna Minnen Beträffande Uppfostran (My Memories of Upbringing) (Castro, de Pablo, Gomez, Arrindell, & Toro, 1997)	EMBU	Not developed in English
47	Egna Minnen Betrffånde Uppfostran for Children (Castro et al., 1997; Markus, Lindhout, Boer, Hoogendijk, & Arrindell, 2003)	EMBU-C	Not a parent-report measure
48	Emotional and Physical Abuse Questionnaire (Kemper, Carlin, & Buntain-Ricklefs, 1994)	EPAB	Not a parent-report measure
49	Environmental Harshness, Health, and Life History Strategy Indicators[b] (Chua, Lukaszewski, Grant, & Sng, 2017)	N/A	Not a measure of child maltreatment
50	Exposure to Community Violence (Richters & Martinez, 1993)	ETV	Not a measure of child maltreatment
51	Exposure to violence questionnaire[b] (Kuo, Mohler, Raudenbush, & Earls, 2000)	N/A	Not a measure of child maltreatment
52	Familial Experiences Questionnaire (Wheelock, Lohr, & Silk, 1997)	FEQ	Not a parent-report measure
53	Family Affective Attitude Rating Scale (Waller, Gardner, Dishion, Shaw, & Wilson, 2012)	FAARS	Not a measure of child maltreatment
54	Family Aggression Screening Tool (Cecil, McCrory, Viding, Holden, & Barker, 2016)	FAST	Not a parent-report measure
55	Family Background Questionnaire-Brief (Melchert & Kalemeera, 2009)	FBQ-B	Not a parent-report measure
56	Family Behaviors Screen (Simmons, Craun, Farrar, & Ray, 2017)	FBS	Not a measure of child maltreatment
57	Family Betrayal Questionnaire (Delker, Smith, Rosenthal, Bernstein, & Freyd, 2017)	FBQ	Not a measure of child maltreatment
58	Family Law Detection of Overall Risk Screen (McIntosh, Wells, & Lee, 2016)	FL-DOORS	Not a measure of child maltreatment
59	Family Maltreatment Diagnostic Criteria (Heyman & Smith Slep, 2009)	N/A	Not a parent-report measure
60	Family Risk of Abuse and Neglect (Lennings, Brummert Lennings, Bussey, & Taylor, 2014)	FRAAN	Not a measure of child maltreatment
61	Family Therapy Alliance Scale (L. N. Johnson, Ketring, & Anderson, 2013)	FTAS	Not a measure of child maltreatment
62	Family Unpredictability Scale (Ross & Hill, 2000)	FUS	Not a measure of child maltreatment
63	Go/No-go Association Task Physical Discipline (Sturge-Apple, Rogge, Peltz, Suor, & Skibo, 2015)	GNAT-Physical Discipline	Not a measure of child maltreatment
64	Home Observation Measure of the Environment (Caldwell & Bradley, 2003)	HOME	Not a parent-report measure
65	Home Safety Screening (Scribano, Stevens, Marshall, Gleason, & Kelleher, 2011)	N/A	Not a measure of child maltreatment
66	Identification of Parents At Risk for Child Abuse and Neglect (van der Put et al., 2017)	IPARAN	Not developed in English
67	Index of Child Care Environment (Anme et al., 2013)	ICCE	Not developed in English
68	Invalidating Childhood Environments Scale (Mountford, Corstorphine, Tomlinson, & Waller, 2007)	ICES	Not a measure of child maltreatment
69	Inventory on Beliefs and Attitudes Towards Domestic Violence (Hutchinson & Doran, 2017)	N/A	Not a measure of child maltreatment
70	ISPCAN Child Abuse Screening Tool Children's Version (Zolotor et al., 2009)	ICAST-C	Not a parent-report measure
71	ISPCAN Child Abuse Screening Tool Parents' Version (Runyan et al., 2009)	ICAST-P	Developed in multiple languages
72	ISPCAN Child Abuse Screening Tools Retrospective Version (Dunne et al., 2009)	ICAST-R	Not a parent-report measure
73	Japanese version of Conflict Tactics Scale[b] (Baba et al., 2017)	CTS1: Japanese version	Developed in English but translated and validated in other languages
74	Juvenile Victimization Questionnaire (Finkelhor, Hamby, Ormrod, & Turner, 2005)	JVQ	Not a parent-report measure

(Continued)

Appendix B. *(continued)*

No	Instrument[a] *(alphabetical order)*	Abbreviation	Reason for exclusion
75	Maternal Characteristics Scale (Polansky, Gaudin, & Kilpatrick, 1992)	MCS	Not a measure of child maltreatment
76	Maternal discipline and appropriateness[b] (Padilla-Walker, 2008)	N/A	Not a parent-report measure
77	Maternal Responsiveness Questionnaire (Leerkes & Qu, 2017)	MRQ	Not a measure of child maltreatment
78	Maternal Self-report Support Questionnaire (D. W. Smith et al., 2010)	MSSQ	Not a measure of child maltreatment
79	Maternal Support Questionnaire–Child Report (D. W. Smith et al., 2017)	MSQ-CR	Not a measure of child maltreatment
80	Meaning of the Child Interview (Grey & Farnfield, 2017)	MotC	Not a measure of child maltreatment
81	Measure of Parenting Style (Parker et al., 1997)	MOPS	Not a parent-report measure
82	Measure Trauma Associated with Child Sexual Abuse (Choudhary, Satapathy, & Sagar, 2018)	MSCSA	Not a measure of child maltreatment
83	Measures of Community-Relevant Outcomes for Violence Prevention Programs[b] (Hausman et al., 2013)	N/A	Not a measure of child maltreatment
84	Medical History Questionnaire[b] (Famularo, Fenton, & Kinscherff, 1992)	N/A	Not a measure of child maltreatment
85	Minnesota Multiphasic Personality Inventory-2 (Butcher, Dahlstrom, Graham, Tellegen, & Kreammer, 1989)	MMPI-2	Not a measure of child maltreatment
86	Multidimensional Assessment of Parenting Scale (Parent & Forehand, 2017)	MAPS	Not a measure of child maltreatment
87	Multidimensional Inventory for Assessment of Parental Functioning (Reis, Orme, Barbera-Stein, & Herz, 1987)	N/A	Not a measure of child maltreatment
88	Multidimensional Neglectful Behavior Scale: Adolescent and Adult Recall Version (Dubowitz et al., 2011)	MNBS-A	Not a parent-report measure
89	Multidimensional Neglectful Behavior Scale-Child Report (Beyazit & Ayhan, 2018)	MNBS-CR	Not a parent-report measure
90	National Council on Crime and Delinquency Indicators (Wood, 1997)	N/A	Not a parent-report measure
91	Needs-Based Assessment of Parental (Guardian) Support (Bolen, Lamb, & Gradante, 2002)	NAPS	Not a measure of child maltreatment
92	Neglect Scale (Harrington, Zuravin, DePanfilis, Ting, & Dubowitz, 2002)	N/A	Not a parent-report measure
93	Parent Cognition Scale[b] (Snarr, Slep, & Grande, 2009)	N/A	Not a measure of child maltreatment
94	Parent discipline style[b] (Mezzich et al., 2007)	N/A	Not a measure of child maltreatment
95	Parent Perception Inventory (Glaser, Horne, & Myers, 1995)	PPI	Not a measure of child maltreatment
96	Parent Perception Inventory-Child version (Bruce et al., 2006)	PPIC	Not a measure of child maltreatment
97	Parent Problem Checklist (Stallman, Morawska, & Sanders, 2009)	PPC	Not a measure of child maltreatment
98	Parent Qualities Measure (Crick, 2006; Stallman et al., 2009)	PQM	Not a measure of child maltreatment
99	Parent Threat Inventory (Crick, 2006; Scher, Stein, Ingram, Malcarne, & McQuaid, 2002)	PTI	Not a parent-report measure
100	Parental Acceptance-Rejection Questionnaire (Rohner & Khaleque, 2005)	PARQ	Not a parent-report measure
101	Parental Anger Inventory (Scher et al., 2002; Sedlar & Hansen, 2001)	PAI	Not a measure of child maltreatment
102	Parental Authority Questionnaire (Buri, 1991)	PAQ	Not a measure of child maltreatment
103	Parental Emotion Regulation Inventory (Lorber, Del Vecchio, Feder, & Smith Slep, 2017; Sedlar & Hansen, 2001)	PERI	Not a measure of child maltreatment
104	Parental Empathy Measure (Kilpatrick, 2005; Lorber et al., 2017)	PEM	Not a measure of child maltreatment
105	Parent-Child Activities Interview (Kilpatrick, 2005; Lefever et al., 2008)	PCA	Not a parent-report measure
106	PARENT-INFANT RELATIONSHIP GLOBAL ASSESSMENT SCALE (Lefever et al., 2008; THREE, 2005)	PIR-GAS	Not a measure of child maltreatment
107	Parenting Anxious Kids Ratings Scale-Parent Report (Flessner, Murphy, Brennan, & D'Auria, 2017; THREE, 2005)	PAKRS-PR	Not a measure of child maltreatment
108	Parenting Behavior Rating Scales (Flessner et al., 2017; G. A. King, Rogers, Walters, & Oldershaw, 1994)	N/A	Not a parent-report measure
109	Parenting Daily Diary (G. A. King et al., 1994; Peterson, Tremblay, Ewigman, & Popkey, 2002)	N/A	Not a parent-report measure
110	Parenting Practices Questionnaire-Corporal Punishment (Avinun, Davidov, Mankuta, Knafo-Noam, & Knafo-Noam, 2018)	PPQ-CP	Not a measure of child maltreatment

(Continued)

Appendix B. *(continued)*

No	Instrument[a] *(alphabetical order)*	Abbreviation	Reason for exclusion
111	Parenting Scale (Peterson et al., 2002; Salari, Terreros, & Sarkadi, 2012)	PS	Not a measure of child maltreatment
112	Parenting Support Needs Assessment (Murry & Lewin, 2014; Salari et al., 2012)	PSNA	Not a measure of child maltreatment
113	Plotkin Child Vignettes (Plotkin, 1983)	PCV	Not a measure of child maltreatment
114	Post-Divorce Parental Conflict Scale (Morris & West, 2000; Murry & Lewin, 2014)	PPCS	Not a measure of child maltreatment
115	Preschool Symptom Self-Report (Martini, Strayhorn, & Puig-Antich, 1990)	PRESS	Not a measure of child maltreatment
116	Production of Discipline Alternatives (Rodriguez, Wittig, & Christl, 2019)	PDA	Not a parent-report measure
117	Protective Factors Survey (Counts, Buffington, Chang-Rios, Rasmussen, & Preacher, 2010; Martini et al., 1990)	PFS	Not a measure of child maltreatment
118	Psychological Maltreatment Rating Scales (Brassard, Hart, & Hardy, 1993; Counts et al., 2010)	PMRS	Not a parent-report measure
119	Psychological Neglect (Brassard et al., 1993; Christ, Kwak, & Lu, 2017)	N/A	Not a parent-report measure
120	Psychologically Violent Parental Practices Inventory (Christ et al., 2017; Gagne, Pouliot-Lapointe, & St-Louis, 2007)	PVPPI	Not developed in English
121	Questionnaire for evaluating maltreatment and neglect (Calheiros, Patrício, Graça, & Magalhães, 2018)	N/A	Not developed in English
122	Reflective Parenting Assessment (Ensink, Leroux, Normandin, Biberdzic, & Fonagy, 2017; Gagne et al., 2007)	RPA	Not a measure of child maltreatment
123	Responsiveness Index (Ensink et al., 2017; Yates, Hull, & Huebner, 1983)	N/A	Not a parent-report measure
124	Revised Child Anxiety and Depression Scale Parent Version (Ebesutani, Tottenham, & Chorpita, 2015; Yates et al., 1983)	RCADS-P	Not a measure of child maltreatment
125	Risk Scale[b] (Ebesutani et al., 2015; Grietens, Geeraert, & Hellinckx, 2004)	N/A	Not a parent-report measure
126	Rorschach Inkblot Method (Choca, 2013; Grietens et al., 2004)	RIM	Not a measure of child maltreatment
127	Scale of Negative Family Interactions (Choca, 2013; Simonelli, Mullis, & Rohde, 2005)	SNFI	Not a parent-report measure
128	Screen for Adolescent Violence Exposure for children version (Flowers, Lanclos, & Kelley, 2002; Simonelli et al., 2005)	KID-SAVE	Not a parent-report measure
129	Sexual Abuse Indicators (Flowers et al., 2002; Terrell et al., 2008)	SAI	Not a parent-report measure
130	Sexual Behavior Problems Questionnaire[b] (Hall, Mathews, & Pearce, 1998; Terrell et al., 2008)	N/A	Not a parent-report measure
131	Sexual Events Questionnaire (Finkelhor, 1979; Hall et al., 1998)	SEQ	Not a parent-report measure
132	Sexual Experiences Survey (Finkelhor, 1979; Koss & Gidycz, 1985)	SES	Not a parent-report measure
133	Shaken Baby Syndrome Awareness Assessment (Koss & Gidycz, 1985; Russell & Britner, 2006)	SBS	Old version of a revised measure
134	Sixteen Personality Factor Questionnaire (Francis, Hughes, & Hitz, 1992; Russell & Britner, 2006)	16-PF	Not a measure of child maltreatment
135	Social Factors and Children Violence Questionnaire (Francis et al., 1992; Oni & Adetoro, 2014)	SPCVQ	No psychometric data found
136	Standardized Observation Codes (Cerezo, Keesler, Dunn, & Wahler, 1986; Oni & Adetoro, 2014)	SOC III	Not a measure of child maltreatment
137	Structured Problem Analysis of Raising Kids (Cerezo et al., 1986; Staal, van den Brink, Hermanns, Schrijvers, & van Stel, 2011)	SPARK	Not a measure of child maltreatment
138	Supervisory Neglect (Coohey, 2003; Staal et al., 2011)	N/A	Not a parent-report measure
139	Symptoms of Trauma Scale (Coohey, 2003; Ford et al., 2017)	SOTS	Not a measure of child maltreatment
140	Trauma Experiences Checklist (Cristofaro et al., 2013; Ford et al., 2017)	TEC	Not a measure of child maltreatment
141	Trauma history questionnaire (Cristofaro et al., 2013; Hooper, Stockton, Krupnick, & Green, 2011)	THQ	Not a parent-report measure
142	Trauma Symptom Checklist for Children (Briere et al., 2001; Hooper et al., 2011)	TSCC	Not a measure of child maltreatment
143	Trauma Symptom Checklist for Young Children (Briere et al., 2001)	TSCYC	Not a measure of child maltreatment

(Continued)

Appendix B. *(continued)*

No	Instrument[a] *(alphabetical order)*	Abbreviation	Reason for exclusion
144	U.S. military's Family Advocacy Program Severity Index (Briere et al., 2001; A. M. Slep & Heyman, 2004)	USAF-FAP Severity Index	Not a parent-report measure
145	Violent Experiences Questionnaire-Revised (A. R. King & Russell, 2017; A. M. Slep & Heyman, 2004)	VEQ-R	Not a parent-report measure
146	Weekly Problems Scales (A. R. King & Russell, 2017; Sawyer, Tsao, Hansen, & Flood, 2006)	WPS	Not a measure of child maltreatment
147	When Bad Things Happen Scale (Fletcher, 1995; Sawyer et al., 2006)	WBTH	Not a measure of child maltreatment
148	Young Parenting Inventory (Young, Klosko, & Weishaar, 2003)	YPI	Not a parent-report measure
149	Young Parenting Inventory-Revised (Louis, Wood, & Lockwood, 2018)	YPI-R2	Not a parent-report measure
150	Young Schema Questionnaire-Short form 3 (Young, 2005)	YSQ-S3	Not a parent-report measure

Notes. N/A = Not Applicable (No Abbreviation).

[a] References of the excluded instruments in this review are available from the first author upon request.

[b] Unofficial title retrieved from publication content as an instrument published without a title or abbreviation.

Appendix C. *Descriptions of the Development and Content Validity Studies on Included Instruments.*

Source[a] *(alphabetical order)*	Instrument	Purpose of study	Study population[b]	Age[c] (range [R] and/or Mean [MN] and/or Standard Deviation [SD])
Bavolek et al. (1979)	**Adult Adolescent Parenting Inventory-2 (AAPI-2)**	To develop and validate the AAPI (as an original version of the AAPI-2)	N = 9 (Stage: Construct development): (I) Professionals in child maltreatment N = 3,000 (Stage: Pilot Testing): (II) Adolescents attending high schools (grade 10-12)	(I) R = NR, MN = NR, SD = NR; (II) R= NR, MN = NR, SD = NR
Gordon et al. (1979)	**Intensity of Parental Punishment Scale (IPPS)**	To develop and validate the IPPS	N = 417: (I) n = 301: Parents of 5- to 10-year-old children; (II) n = 50: Upper-middle-class parents of 7- to 12-year old children; (III) n = 26: Mothers of 6- to 9-year-old children; (IV) n = 40: Mothers of 6- to 14-year-old children	(I) R = NR, MN = NR, SD = NR; (II) R = NR, MN = NR, SD = NR; (III) R= NR, MN = NR, SD = NR; (IV) R = NR, MN = NR, SD = NR
Heyman et al. (2019)	**Family Maltreatment-Child Abuse criteria (FM-CA)**	To develop and validate the FM-CA	N = 126: U.S. Air Force service members and their spouses (F = 41; M = 85)	R = NR, MN = NR, SD = NR
Holden and Zambarano (1992)	**Parental Response to Child Misbehavior questionnaire (PRCM)**	To exam parental responses to children's misbehavior in maternal reported use of physical punishment by using the CPSS and the PRCM	N = 132: Mothers of 12- to 48-month-old children (F = 132; M = 0)	R = 20-44y, MN = 31.4y, SD = 4.5y
Lang and Connell (2017)	**Child Trauma Screen-Exposure Score (CTS-ES)**	To develop and validate the CTS-ES	N = 923 (Stage: CTS-ES Development): (I) Parents of children receiving care at outpatient behavioral health clinics N = 69 (Stage: CTS-ES Validation): (II) Parents of children receiving care at outpatient behavioral health clinics	(I) R = NR, MN = NR, SD = NR; (II) R = NR, MN = NR, SD = NR
Loeber et al. (1998)	**Child Neglect Scales-Maternal Monitoring and Supervision scale (CNS-MMS)**	To examine delinquency, substance use, early sexual behavior, and mental health problems of urban boys by using diverse instruments including the SIS (as an original version of the CNS-MMS)	N = 1507: (I) n = 503: parents with boys in the first grade in Pittsburgh public schools; (II) n = 508: parents with boys in the fourth grade in Pittsburgh public schools (III) n = 506: parents with boys in the seventh grade in Pittsburgh	(I) R = NR, MN = NR, SD = NR; (II) R = NR, MN = NR, SD = NR; (III) R = NR, MN = NR, SD = NR
Meinch et al. (2018)	**ISPCAN Child Abuse Screening Tool for use in Trials (ICAST-Trial)**	To develop and validate the ICAST-Trial	N = 115 (Stage: Pilot study) (I) Parents of adolescents participated in a parenting program to prevent child abuse (F = 112; M = 3) N = 552 (Stage: Validation of ICAST-Trial) (II) Parents of adolescents participated in a parenting program to prevent child abuse (F = 523; M = 29)	(I) R = NR, MN = 48y, SD = 13.6y; (II) R = NR, MN = 49.4y, SD = 14.69y
Runyan et al. (2009)	**ISPCAN Child Abuse Screening Tool for use in Trials (ICAST-Trial)**	To develop and validate the ICAST-P (as an original version of the ICAST-Trial)	N = 51 (Stage: Item development): (I) Professionals in child maltreatment N = 697 (Stage: Pilot Testing): (II) Parents with children under the age of 18 in six different countries	(I) R = NR, MN = NR, SD = NR; (II) R = NR, MN = NR, SD = NR

(continued)

Appendix C. *(continued)*

Source[a] *(alphabetical order)*	Instrument	Purpose of study	Study population[b]	Age[c] (range [R] and/or Mean [MN] and/or Standard Deviation [SD])
Russell and Britner (2006)	**Shaken Baby Syndrome awareness assessment-Short Version (SBS-SV)**	To develop and evaluate the psychometric properties of the SBS (as an original version of the SBS-SV)	N = 288 (Stage: Pilot study) (I) Undergraduate psychology students (F = 207; M = 81) N = 264 (Stage: Validation of SBS) (II) Caregivers and non-caregivers over the age of 18 (F = 191; M = 73)	(I) R = 17-31y, MN = 19y, SD = NR; (II) R = 18-78y, MN = 32y, SD = NR
Straus et al. (1995)	**Mother-Child Neglect Scale (MCNS)**	To describe the development and validation of the MNBS (as an original version of the MCNS)	N = 359: Adolescences and adults (F = 236, M = 123)	R = NR, MN = NR, SD = NR
Straus et al. (1995)	**Mother-Child Neglect Scale-Short Form (MCNS-SF)**	To describe the development and validation of the MNBS-SF (as an original version of the MCNS-SF)	N = 359: Adolescences and adults (F = 236, M = 123)	R = NR, MN = NR, SD = NR
Straus et al. (1998)	**Conflict Tactics Scales: Parent-Child version (CTSPC)**	To develop and test the reliability and validity of CTSPC	N = 1,000: Parents of children under 18 years old participated in an U.S. national survey (F = 660; M = 340)	R = NR, MN = 36.8y, SD = NR
Stewart et al. (2015)	**Child Neglect Questionnaire (CNQ)**	To develop and evaluate psychometric properties of the CNQ	N = 172: (I) n = 76: Parents of children having fathers with Substance Use Disorder (SUD); (II) n = 96: Parents of children having fathers without SUD	(I) R = NR, MN = NR, SD = NR; (II) R = NR, MN = NR, SD = NR
Twentyman et al. (1981)	**Parent Opinion Questionnaire (POQ)**	To develop and validate the POQ	N = 30 (Stage: Item development): (I) n = 23: Child protective case workers (II) n = 7: Health nurses N = 15 (Stage: Cross validation): (III) Child protective case workers	(I) R = NR, MN = NR, SD = NR; (II) R = NR, MN = NR, SD = NR; (III) R = NR, MN = NR, SD = NR
Zaidi et al. (1989)	**Analog Parenting Task (APT)**	To determine whether there was an association between punitive childhood histories by the AEIII and abusive parenting by the APT.	N = 86 (Stage: preliminary study) (I) n= 49: university students experienced severe physical punishment in childhood (F = 19; M = 30); (II) n= 37: university students experienced mild physical punishment in childhood (F = 26; M = 11) N = 338 (Stage: main study) (III) n = 169: Mothers of children referred for child psychiatry service (F = 169; M = 0); (IV) n = 169: Fathers of children referred for child psychiatry service (F = 0; M = 169)	(I) R = 18-24y, MN = 19.4y, SD = NR; (II): R = 17-23y, MN = 19.0y, SD = NR; (III) R = 22-51y, MN = 34.2y, SD = NR; (IV) R = 22-57y, MN = 36.8y, SD = NR

Notes. AAPI = Adult Adolescent Parenting Inventory; AEIII = Assessing Environments III; CAP = Child Abuse Potential inventory; CPSS = Computer-Presented Social Situations; ICAST-P = ISPCAN (International Society for the Prevention of Child Abuse and Neglect) Child Abuse Screening Tool-Parent version; MNBS = Multidimensional Neglectful Behavior Scale; MNBS-SF = Multidimensional Neglectful Behavior Scale-Short Form; SIS = Supervision and Involvement Scale; SBS = Shaken Baby Syndrome awareness assessment.

[a] References of the development and content validity studies on included instruments can be found in the reference section of this review.

[b] N = total sample size; n = subgroups; M = male; F = female.

[c] R = range; MN = mean; Med = median; NR = not reported; SD = standard deviation; NR = Not Reported.